

UNIVERSIDADE FEDERAL DO PARANÁ

CARLOS EDUARDO IBALDO GONÇALVES

ANÁLISE DE VARIANTES EM GENES HUMANOS QUE REGULAM A
PIGMENTAÇÃO PARA PREDIÇÃO DE FENÓTIPOS DE COR DE OLHOS E DE
PELE PARA FINS FORENSES

CURITIBA

2019

CARLOS EDUARDO IBALDO GONÇALVES

ANÁLISE DE VARIANTES EM GENES HUMANOS QUE REGULAM A
PIGMENTAÇÃO PARA PREDIÇÃO DE FENÓTIPOS DE COR DE OLHOS E DE
PELE PARA FINS FORENSES

Tese entregue à banca avaliadora para obtenção
do título de Doutor pelo Programa de Pós-
Graduação em Genética do Setor de Ciências
Biológicas da Universidade Federal do Paraná.

Orientadora: Profa. Dra. Maria da Graça Bicalho

Coorientadores: Profa. Dra. Clarice Sampaio Alho
(PUCRS) e Perito Dr. Marcelo Malaghini (Instituto
de Criminalística do Paraná)

CURITIBA

2019

Universidade Federal do Paraná. Sistema de Bibliotecas.
Biblioteca de Ciências Biológicas.
(Dulce Maria Bieniara – CRB/9-931)

Gonçalves, Carlos Eduardo Ibaldo

Análise de variantes em genes humanos que regulam a pigmentação
para predição de fenótipos de cor de olhos e de pele para fins forenses. /
Carlos Eduardo Ibaldo Gonçalves. – Curitiba, 2019.
273 p.: il.

Orientadora: Maria da Graça Bicalho

Coorientadora: Clarice Sampaio Alho

Coorientador: Marcelo Malaghini

Tese (doutorado) - Universidade Federal do Paraná, Setor de Ciências
Biológicas. Programa de Pós-Graduação em Genética.

1. Genética forense 2. Pigmentação da pele 3. Fenótipo 4.
Polimorfismo de nucleotídeo único 5. Cor de olho I. Título II. Bicalho,
Maria da Graça III. Alho, Clarice Sampaio IV. Malaghini, Marcelo V.
Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa
de Pós-Graduação em Genética.

CDD (20. ed.) 575.12



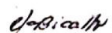
MINISTÉRIO DA EDUCAÇÃO
SETOR SETOR DE CIÊNCIAS BIOLÓGICAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO GENÉTICA -
40001016006P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GENÉTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **CARLOS EDUARDO IBALDO GONCALVES** intitulada: **Análise de variantes em genes humanos que regulam a pigmentação para predição de fenótipos de cor de olhos e de pele para fins forenses.**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua Aprovação no rito de defesa.

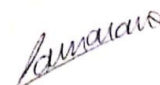
A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 15 de Março de 2019.


MARIA DA GRAÇA BICALHO
Presidente da Banca Examinadora


ALEJANDRO CORREA DOMÍNGUEZ
Avaliador Externo (ICC/FIOCRUZ)


MARCIA HOLSBACH BELTRAME
Avaliador Interno (UFPR)


LEONARDO ARDUINO MARANO
Avaliador Externo (PC-PR)


DANIELLE M. FERREIRA
Avaliador Interno (UFPR)

A Deus, à minha família, aos meus orientadores, aos meus colegas e amigos que contribuíram para a realização deste trabalho. Todos fazem parte do que sou hoje.

AGRADECIMENTOS

Às professoras Maria da Graça Bicalho e Clarice Sampaio Alho pela orientação, pelo apoio, pelo incentivo, pelo aprendizado, pela amizade e pela confiança. Ao perito criminal Marcelo Malaghini pela coorientação. Foi uma honra poder trabalhar e conviver com vocês durante esses anos, levarei esses momentos com muita alegria por toda a minha vida!

Aos professores Alejandro Correa Dominguez, Danielle Malheiros Ferreira, Márcia Holsbach Beltrame e ao perito criminal Leonardo Arduino Marano pela oportunidade de aprimorar a minha tese com os seus conhecimentos e conselhos.

À toda equipe LIGH pelo aprendizado, pelas boas risadas e pela ótima convivência desde o mestrado e que acompanhou o desenvolvimento deste projeto, mesmo à distância, durante o doutorado.

À toda equipe do LGHM da PUCRS, pessoas incríveis que estarão para sempre no meu coração, amo todos vocês! Viva a República da Chuvashia! Go Tigers!

Ao professor da UFRGS e amigo Márcio Dorn, sem a tua ajuda e a tua expertise na bioinformática este trabalho não existiria.

Ao Alessandro Kahmann pelo auxílio nas estatísticas.

Ao Programa de Pós-Graduação em Genética da Universidade Federal do Paraná e ao Programa de Pós-Graduação em Biologia Celular e Molecular da Pontifícia Universidade Católica do Rio Grande do Sul, pela oportunidade de aprimorar meus conhecimentos em um ensino de alto nível.

À CAPES (Edital 25/2014 - Pró-Forenses), ao CNPq (Edital INCT) e à FAPERGS pelo apoio na realização deste trabalho.

À CAPES pela bolsa concedida nestes quatro anos de trabalho.

A todos os professores e colegas, seriam muitos a citar, que contribuíram na minha formação.

À minha família, principalmente aos meus pais, Assis e Marta, sem eles nada teria sido possível! Muito obrigado pelo apoio, pela atenção, pela dedicação, pelo carinho e pelo amor! Amo vocês!

À minha irmã Aline e ao meu cunhado Roben pela amizade e pelo apoio.

À minha vó Marina, *in memoriam*. Saudades, morena!

Ao meu irmão canino Caco e aos meus sobrinhos caninos Gene e Twix, que tornam as nossas vidas muito mais felizes.

Aos meus amigos, seriam muitos a citar, pela amizade, pela convivência, pelo incentivo e pelo apoio moral e psicológico em diversas circunstâncias da vida.

A todas as dificuldades, que por ventura tenham trazido aprendizados.

Aos meus orientadores espirituais e, especialmente, ao meu Mentor, pela confiança e auxílio de sempre.

A Deus, essa energia universal, presente em tudo o que existe e que nos une.

“A persistência é o caminho do êxito”.

Charles Chaplin

RESUMO

Variantes genéticas relacionadas a características externamente visíveis são de grande relevância para a predição da cor de olhos e de pele em humanos. Neste estudo, investigamos a capacidade de um conjunto de SNPs de genes relacionados à pigmentação humana para prever a cor de olhos e de pele em uma população do sul do Brasil. Em uma primeira análise, investigamos um conjunto de oito SNPs de sete genes, em 478 indivíduos do sul do Brasil com diferentes fenótipos de pele e de olhos, genotipados através do Sistema Multiplex SNaPshot®. Então, usando regressão logística multinomial, criamos cinco modelos para predição do fenótipo de olhos: A (5 classes: Azul, Verde, Mel, Castanho Claro e Castanho Escuro, acurácia geral 75%); B (3 classes: Azul, Intermediário e Castanho Escuro, 82%); C (2 classes: Azul e Não Azul, 94%); D (2 classes: Castanho e Não Castanho, 91%); e E (2 classes: Claro e Não Claro, 95%). Além disso, criamos três modelos para predição de fenótipo de pele: F (6 classes: Branco, Pálido, Bege, Marrom Claro, Marrom Médio e Marrom Escuro, 49%); G (3 classes: Claro, Intermediário e Escuro, 82%); e H (2 classes: Claro e Escuro, 92%). Em uma segunda análise, investigamos um conjunto de 68 SNPs de 21 genes, em 465 indivíduos do sul do Brasil com diferentes fenótipos de pele e de olhos, genotipados no Sistema MassARRAY®, utilizando reagentes iPLEX® Pro. Então, usando ferramentas de bioinformática, foram analisados vários modelos para predição de fenótipo de olhos e de pele e, a partir deles, foi escolhido o melhor de cada característica para construir O BR-EASE (BRAZILIAN EYE AND SKIN ESTIMATOR), um sistema fácil de fenotipagem forense através do DNA. O sistema classifica os olhos em três classes (azul, intermediário e castanho) com uma acurácia geral de 87%, classificando corretamente 98% de olhos castanhos, 94% de olhos azuis e 31% de olhos com fenótipo intermediário. A pele também é classificada em três classes (clara, intermediária, escura) e apresenta uma acurácia geral de 97,5%, classificando corretamente 100% de pele clara, 98% de pele escura e 90% de pele intermediária. Esses modelos foram capazes de prever a pigmentação do olho e da pele, com alta acurácia em indivíduos com fenótipo claro ou escuro, e menor em indivíduos de fenótipo intermediário. É importante que os órgãos policiais possuam preditores de fenótipos baseados na população local para auxiliar em suas investigações, principalmente porque a população brasileira é de ancestralidade

miscigenada. As análises apresentadas aqui são uma contribuição importante para a área de fenotipagem de DNA para fins forenses.

PALAVRAS-CHAVE: BR-EASE. Forense. Pigmentação Humana. Cor dos olhos. Cor da pele. Fenotipagem de DNA. SNP.

ABSTRACT

Genetic variants related to externally visible characteristics are of great relevance for the prediction of eye and skin colors in humans. In this study, we tested the ability of a set of SNPs in pigment-related genes to predict eye and skin colors in an admixed South Brazilian population. In a first analysis, we investigated a set of eight SNPs of seven genes, in 478 Southern Brazilian subjects with diverse skin and eye colors, genotyped by SNaPshot® Multiplex System. Then, using multinomial logistic regression, we designed five models for eye phenotype prediction: A (5 classes: Blue, Green, Hazel, Light and Dark Brown, general accuracy 75%); B (3 Classes: Blue, Intermediate and Dark Brown, 82%); C (2 classes: Blue and Non-Blue, 94%); D (2 classes: Brown and Non-Brown, 91%); and E (2 classes: Light and Non-Light, 95%). Also, we designed three models for skin phenotype prediction: F (6 classes: White, Pale, Beige, Light Brown, Medium Brown and Dark Brown, 49%); G (3 classes: Light, Intermediate and Dark, 82%); and H (2 classes: Light and Dark, 92%). In a second analysis, we investigated a set of 68 SNPs of 21 genes, in 465 Southern Brazilian subjects with various skin and eye colors, genotyped by iPLEX® Pro reagents on the MassARRAY® System. Then, using bioinformatics tools, we analyzed several models for eyes and skin phenotype prediction. We chose the best model of each feature to construct BR-EASE (BRAZILIAN EYE AND SKIN ESTIMATOR), an easy system of forensic phenotyping through DNA. This system classifies eyes into three classes (blue, intermediate and brown) with an overall accuracy of 87%, correctly classifying 98% of brown eyes, 94% of blue eyes and 31% of eyes with intermediate phenotype. Likewise, skin is classified into three classes (light, intermediate, dark) and has an overall accuracy of 97.5%, correctly classifying 100% of light skin, 98% of dark skin and 90% of intermediate skin. These models were able to predict eye and skin pigmentation, with high assertiveness in subjects with light or dark phenotype, and lower in subjects of intermediate phenotype. It is important for law enforcement agencies to have phenotype predictors based on local population to assist in their investigations, mainly because the Brazilian population is of admixed ancestry. Analyzes presented here are an important contribution to the area of forensic DNA phenotyping.

KEYWORDS: BR-EASE. Forensic. Human Pigmentation. Eyes Color. Skin Color.
DNA Phenotyping. SNP.

LISTA DE FIGURAS

FIGURA 1: Produção e distribuição de melanina em peles escuras e claras.....	28
FIGURA 2: Esquema de um melanócito. O mecanismo para a produção da melanina.....	29
FIGURA 3: Pigmentação dos olhos. Diferenças entre os tipos e quantidades de melanossomos nos melanócitos dos olhos mais claros aos mais escuros.....	30
FIGURA 4: Epitélio pigmentoso da íris e refletância de ondas luminosas.....	31
FIGURA 5: Tonalidades de pigmentação de íris encontradas em humanos.....	32
FIGURA 6: Via biossintética da produção de eumelanina/feomelanina a partir de tirosina.....	34
FIGURA 7: Modelo de determinação da cor de Iris castanho (A) e azul (B) influenciado pelo SNP rs12913832 do gene <i>HERC2</i>	36
FIGURA 8: Indicação na estrutura celular dos genes e/ou seus produtos proteicos candidatos para predição de pigmentação em humanos.....	39

FIGURA 9: Modelos de predição de pigmentação de olhos.....	108
FIGURA 10: Modelos de predição de pigmentação da pele.....	109
FIGURA 11: BR-EASE – Estimador de cor de olhos.....	110
FIGURA 12: BR-EASE – Estimador de cor de pele.....	111

LISTA DE QUADROS

QUADRO 1: Genes e SNPs candidatos à predição de pigmentação em humanos.....	38
--	-----------

QUADRO 2: Outros genes e SNPs candidatos à predição de pigmentação em humanos.....	45
---	-----------

LISTA DE ABREVIATURAS

ABO: ABO Blood System
ACC: Accuracy
AIC: Akaike Information Criterion
AIM1: Absent in Melanoma 1 Protein
Alpha-MSH (α -MSH): Alpha-Melanocyte-Stimulating Hormone
BR-EASE: Brazilian eye and skin estimator
cAMP: Cyclic Adenosine Monophosphate
CD: DOPAcisteinila
CEV: Características Externas Visíveis
X2: Qui Quadrado
CREB: CyclicAMP Response-Element Binding Protein
DCT: Dopacromo Tautomerase
DHI: 5,6-Dihidroindole
DHICA: 5,6-Dihidroindole-2-Carboxílico
DNA: Deoxyribonucleic Acid
DOPA: 3,4-Dihidroxifenilalanina
DQ: Dopaquinona
EVC: Externally Visible Characteristics
FDP: Forensic DNA Phenotyping
GWAS: Genome-Wide Association Study
HERC2: HECT Domain and RCC1-Like Domain-Containing Protein 2
HLA: Human Leukocyte Antigen
HLTF: Helicase-Like Transcription Factor
Kb: Kilobase
LEF1: Lymphoid Enhancer-Binding Factor 1
MATP: Membrane-Associated Transporter Protein
MC1R: Melanocortin 1 Receptor
MITF: Microphthalmia-associated Transcription Factor
MLP: Multi-Layer Perceptron
MN: MN Blood System
mtDNA: DNA Mitochondrial
NCKX5: Na⁺/Ca²⁺/K⁺ Exchanger 5

ng: Nanograma
NGS: Next-Generation Sequencing
OCA2: Oculocutaneous Albinism Type II
PCR: Polymerase Chain Reaction
pg: Picograma
PMEL17: Melanosomal Protein 17
RF: Random Forest
RFE: Recursive Feature Elimination
RGB: Red, Green, Blue
Rh: Rh Blood System
SAP: Shrimp Alkaline Phosphatase
SLC24A5: Solute Carrier Family 24, Member 5
SLC45A2: Solute Carrier Family 45, Member 2
SNP: Single Nucleotide Polymorphism
STR: Short Tandem Repeat
SVM: Support Vector Machine
TYR: Tyrosinase
TYRP1: Tyrosinase-Related Protein 1
USA: United States of America
UV: Ultraviolet

SUMÁRIO

1 INTRODUÇÃO.....	19
1.1 OBJETIVO.....	20
1.1.1 Objetivos Específicos.....	20
1.2 JUSTIFICATIVA.....	20
2 REVISÃO BIBLIOGRÁFICA.....	21
2.1 PERFIL DE DNA HUMANO NA PERÍCIA FORENSE.....	21
2.2 DNA DE AMOSTRAS DEGRADADAS, ANTIGAS E DE MISTURAS...	22
2.3 ANÁLISE DE SNPS NA GENÉTICA FORENSE.....	25
2.4 PREDIÇÃO DE FENÓTIPO DE PIGMENTAÇÃO HUMANA.....	27
2.4.1 Biologia da Pigmentação.....	27
2.4.1.1 Pigmentação na Pele.....	28
2.4.1.2 Pigmentação no Olho.....	30
2.4.1.3 Melanogênese.....	33
2.5 PIGMENTAÇÃO E ANCESTRALIDADE.....	34
2.6 GENES E SNPS ENVOLVIDOS NA DETERMINAÇÃO DA PIGMENTAÇÃO.....	37
2.6.1 Gene <i>HERC2</i>	39
2.6.2 Gene <i>OCA2</i>	40
2.6.3 Gene <i>SLC24A5</i>	41
2.6.4 Gene <i>SLC45A2</i>	42
2.6.5 Gene <i>TYR</i>	43

2.6.6 Gene <i>TYRP1</i>	44
2.6.7 Gene <i>MC1R</i>	44
2.6.8 Outros Genes e SNPs relacionados à Pigmentação Humana.....	45
3 RESULTADOS E DISCUSSÃO	49
3.1 ARTIGO I.....	50
3.2 ARTIGO II.....	74
4 DISCUSSÃO GERAL E CONCLUSÃO	107
REFERÊNCIAS	113
ANEXOS	123
I MATERIAIS E MÉTODOS.....	123
II BR-EASE SUPPLEMENTARY MATERIAL.....	124
III MATERIAIS INFORMATIVOS DO PROJETO.....	268
IV FICHA DE CADASTRO.....	272
V TERMO DE CONSENTIMENTO.....	273
VI PARECER DO COMITÊ DE ÉTICA.....	274

1 INTRODUÇÃO

Homicídios, agressão sexual, crimes contra patrimônio, vítimas desaparecidas, sepultamentos coletivos, acidentes massivos são itens presentes na rotina de trabalho dos peritos forenses. Análises de balística, papiloscopia, toxicologia, perícias em vídeos/fono/áudio, medicina legal, análises históricas, levantamento de testemunhas são exemplos de ferramentas que contribuem para a investigação dos fatos e para a identificação de vítimas, envolvidos e culpados. Como estas ferramentas, as avaliações de DNA de amostras relacionadas aos elementos de cada caso têm sido também amplamente utilizadas. Agora, com a disponibilidade das novas tecnologias para sequenciamento e genotipagem, a utilização do DNA se estende para resolver casos antes impossibilitados pelos limites tecnológicos. Contar com estas novas tecnologias será crucial para elucidar casos, em especial, aqueles difíceis em que as ferramentas padrões para análise de DNA em identificação humana mostram-se incapazes de gerar resultados.

A fenotipagem forense pelo DNA se apresenta como uma abordagem promissora para suprir lacunas na busca de pessoas desconhecidas, em investigações criminais, e na identificação de vítimas de catástrofes e de pessoas desaparecidas. Essa metodologia permite a previsão individual de características externamente visíveis (CEV) a partir de análises de polimorfismos de nucleotídeo único (SNPs) informativos de fenótipos. Entre esses SNPs, os mais bem descritos são aqueles relacionados com as características de pigmentação, como cor dos olhos, pele e cabelo. Estudos vêm demonstrando o elevado poder de predição dessas CEV, apresentando resultados satisfatórios na predição da cor de íris azul e castanha e de cabelo ruivo, enquanto para as demais ainda são necessárias mais pesquisas para prever com precisão esses fenótipos. Ainda, devido à maioria dos marcadores descritos serem relacionados às populações europeias, são necessários estudos em outras populações, entre elas a do Brasil, que é extremamente miscigenada. Em um futuro próximo essa metodologia estará apta a integrar às rotinas forenses, com grande aplicabilidade e confiabilidade.

Neste trabalho, realizamos estudos para construir modelos de predição de fenótipo de olhos e de pele para fins forenses, a partir da análise de 68 SNPs de 19 genes relacionados à pigmentação humana em indivíduos do sul do Brasil.

1.1 OBJETIVO

Construir modelos genéticos de predição de fenótipo de cor de olhos e de pele para fins forenses.

1.1.1 Objetivos Específicos

- Identificar as possíveis associações estatísticas entre fenótipos e alelos de genes envolvidos na síntese de melanina.
- Investigar um conjunto de sete SNPs (de sete genes) e sua eficiência na predição de cor de olhos e de pele, em 478 indivíduos com diferentes fenótipos do sul do Brasil, utilizando modelos de regressão logística multinomial.
- Investigar um conjunto de sessenta e oito SNPs (de vinte e um genes) e sua eficiência na predição de cor de olhos e de pele, em 465 indivíduos com diferentes fenótipos do sul do Brasil, utilizando técnicas de bioinformática, sendo duas técnicas de seleção de características: RFE (Recursive Feature Elimination) e X^2 (Teste do Qui Quadrado); e três algoritmos de aprendizado de máquina: SVM (Support Vector Machine), RF (Random Forest) e MLP (Multi-Layer Perceptron).
- Estimar a frequência alélica e genotípica dos 68 SNPs numa população miscigenada do sul do Brasil.

1.2 JUSTIFICATIVA

A fenotipagem forense pelo DNA vem se destacando como uma poderosa ferramenta de auxílio nas investigações criminais. Entretanto, ainda são necessários estudos adicionais e pesquisas por novos SNPs, que predigam fenótipos com a maior fidelidade possível, para aumentar a precisão das metodologias e fornecer alta confiabilidade para uso forense.

2 REVISÃO BIBLIOGRÁFICA

2.1 PERFIL DE DNA HUMANO NA PERÍCIA FORENSE

A utilização do DNA como ferramenta de identificação humana veio com a publicação de Jeffreys e colaboradores, relatando que certas regiões do DNA produziam uma espécie de impressão digital do DNA (Jeffreys *et al.*, 1985). Desde este importante fato a Ciência Forense tem sido beneficiada cada vez mais com a tecnologia de análise do DNA, solucionando questões cruciais. Há cerca de 20 anos, a genotipagem do DNA é utilizada no mundo inteiro para solucionar casos de identificação em crimes violentos, crimes menores, atos de terrorismo, desaparecimento de pessoas e desastres em massa. A possibilidade de análise do DNA de cada indivíduo aumentou a confiabilidade dos resultados de identificação pessoal e de testes de paternidade, pois, ao contrário dos testes sorológicos anteriores (ABO, MN, Rh, HLA, entre outros) que necessitavam de uma fonte de fluido ou sangue, os testes que utilizam DNA não dependem da natureza do material analisado, pois a informação genética está contida integralmente em todas as células somáticas de um indivíduo (Kobachuk, 2012). Outra importante vantagem da análise do DNA é que este é fisicamente mais resistente que as proteínas, pois a informação genética é encontrada na sequência nucleotídica tornando sua análise independente da transcrição.

O principal objetivo da identificação humana na prática forense criminal é prover dados que justifiquem se determinado indivíduo é fonte doadora de uma evidência biológica. Dentre os possíveis resultados da investigação constam: 1- exclusão, quando as amostras biológicas possuem origens diferentes; 2- resultados inconclusivos, quando não é possível determinar se as amostras provêm ou não do mesmo doador; e 3- inclusão quando as amostras são similares e originaram-se da mesma fonte (Inman *et al.*, 2002). A fim de alcançar plenamente este objetivo, ao longo dos anos foram implantadas diferentes técnicas a fim de caracterizar e individualizar pessoas, como o uso dos conjuntos de marcadores autossômicos, altamente polimórficos, conhecidos como os STRs (*short tandem repeats*). Na atualidade, os STRs são os marcadores padrões no sistema forense, sendo aplicado nos casos de identificação forense criminal e em testes de paternidade ou parentesco

(Yoshida *et al.*, 2011). A confiabilidade elevada do resultado de uma análise de STR permite que ele seja usado como prova respeitável nos processos penal ou cível.

Um sistema igualmente eficaz à identificação de variações de STRs para a identificação humana é a análise conjunta de marcadores moleculares do tipo SNPs (*single nucleotide polymorphism*), os quais, da mesma forma, permitem a obtenção de perfis genômicos únicos para cada indivíduo (Kobachuk, 2012).

Uma derivação da análise de DNA para a perícia forense que se mostra agora também muito promissor para auxiliar a investigação policial categorizando vítimas e/ou suspeitos, é a predição de características externas visíveis (CEV), como a predição de grau de pigmentação de pele e de olho. Estudos recentes têm sustentado que a análise de genes diretamente envolvidos no processo de síntese dos pigmentos tem um elevado poder preditivo para inferir fenótipo de coloração de pele e olhos (Liu *et al.*, 2013). Na pele, em maior ou menor quantidade, a melanina é produzida nos melanócitos (camada basal da epiderme) e depositada nos queratinócitos (onde impede lesões no DNA causadas pela luz ultravioleta). Nos olhos, a melanina é produzida pelos melanócitos do epitélio pigmentar da íris, depositando-se no estroma da íris cujas variações de densidade resultam também nas diferenças de cor. Avaliando amostras de indivíduos com pele/olhos claros ou pele/olhos escuros, em populações de ancestralidade europeias e africanas, respectivamente, se têm encontrado associações positivas entre a herança e a manifestação da pigmentação. Estudos associativos entre genótipo e o fenótipo de coloração, contudo, devem conseguir isolar o componente genético associado à ancestralidade para determinar a implicação efetiva no fenótipo de cada variante alélica como geradora de fenótipo claro ou escuro (Liu *et al.*, 2013).

2.2 DNA DE AMOSTRAS DEGRADADAS, ANTIGAS E DE MISTURAS

Os efeitos da degradação ambiental e química têm sido extensivamente investigados em estudos de validação de sistemas em DNA humano. A possibilidade de obtenção de perfis genéticos completos é inversamente proporcional ao tempo em que a amostra fica exposta à atividade de nucleases ou a elementos, agentes e processos que degradam o DNA como fogo, salinidade, processos oxidativos e contaminantes (Witcum *et al.*, 2013). Desastres em massa como os ocorridos no

World Trade Center em 2001 nos Estados Unidos (Biesecker *et al.*, 2005), Tsunami no Oceano Índico em 2004, o incêndio ocorrido em 2004 no supermercado Ycuá Bolaños em Assunção-Paraguai, o furacão Katrina em 2005, ou ainda o acidente ocorrido em 2007 com o voo TAM JJ 3054, no Brasil, são um desafio para os médicos, odontologistas, papiloscopistas e para os setores de biologia molecular encarregados da identificação das vítimas e dos restos mortais. A identificação de corpos humanos em avançado estado de decomposição como, por exemplo, exumados, afogados, esquartejados, ocultados, expostos ao tempo, a ação de animais ou carbonizados fica extremamente prejudicada (Bilge *et al.*, 2003). Isto ocorre porque os elementos utilizados pelos médico-legistas, odonto-legistas e/ou papiloscopistas para tal identificação, como impressões digitais, sexo, compleição física, grupo étnico, estatura e arcada dentária podem estar alterados a ponto de impossibilitar qualquer conclusão. Desta forma, a única possibilidade de identificação é através do uso das técnicas de DNA (Graw *et al.*, 2000). Em muitas situações, a amostra de DNA recolhida na cena do crime, por exemplo, não está somente degradada, mas em concentrações muito baixas (Lowe *et al.*, 2003). Muitos esforços estão sendo feitos no sentido de melhorar a sensibilidade dos processos de análises de DNA; “sensibilidade” aqui é definida como a menor concentração de DNA molde que pode produzir um perfil genético completo. Painéis comerciais convencionais que usam análise de STRs multiplex têm, em média, um limite inferior de sensibilidade de 250pg e uma eficiência ótima quando 1ng de DNA é amplificado. O substrato/suporte onde a amostra forense é encontrada pode também afetar a qualidade da amostra de DNA acelerando sua degradação ou contaminando-o. Por exemplo, há muito se conhece que pigmentos e corantes presentes em brim ou couro, que inibem a análise do DNA, podem ser co-extraídos com a molécula.

Em amostras antigas/ancestrais há um investimento para que se obtenha DNA a partir de ossos e/ou dentes a fim de se contornar a presença de muitas substâncias inibidoras e a baixa quantidade de DNA. Uma vez que os métodos utilizados para extração de DNA a partir de restos de esqueletos antigos também têm um forte efeito sobre o sucesso de amplificação, é importante o uso de um procedimento de extração eficiente capaz de remover todos os inibidores possíveis para se obter o máximo de DNA de boa qualidade disponível. Uma experiência bem sucedida foi a identificação das vítimas da Segunda Guerra Mundial, pois além dos polimorfismos do DNA mitocondrial (mtDNA; que são os mais facilmente analisáveis), foram avaliadas com

sucesso variações de DNA nuclear autossômico e de cromossomo Y (Zupanič Pajnič, 2013). A identificação genética de Leopoldo III, na Áustria, e de cinco esqueletos datados com 300 anos encontrados no túmulo de Auersperg, Ljubljana, Slovenia, foram também realizadas com sucesso através de marcadores autossômicos, além dos mitocondriais (Bauer *et al.*, 2013). Porém, estudos paleogenéticos recentes realizados nos Estados do Rio de Janeiro e do Maranhão, Brasil, demonstraram que foi possível apenas obter sucesso com a análise do mtDNA (região HVS-I) de amostras de ossos da costela dos esqueletos humanos (Marinho *et al.*, 2006; Leite *et al.*, 2014).

Amostras com misturas de DNA de diferentes contribuintes são frequentes nos casos forenses, porém a grande maioria delas se concentra nas amostras recolhidas de crimes de agressão sexual. Estima-se que no Brasil em aproximadamente 25% das ocorrências destes crimes contra mulheres o agressor é um indivíduo desconhecido à vítima (SINAN/SVS/MS, 2013), sendo esse um dos maiores desafios na identificação dos criminosos durante a investigação policial de estupro (Corovic *et al.*, 2012). Outro desafio que a investigação enfrenta é a atuação serial de alguns estupradores, podendo o número de vítimas de um mesmo criminoso variar de duas a até dezenas de vítimas (Woodhams e Labuschagne, 2012). Além disso, estudos indicam que estupradores seriais tendem a abordar mais frequentemente vítimas desconhecidas (LeBeau, 1987). Em casos de agressão sexual, fluidos biológicos como sêmen, saliva e outras secreções podem ser colhidas do corpo da vítima, bem como de preservativos, lençóis, toalhas e roupas deixados no local do crime e, posteriormente, analisadas em laboratório, a fim de se identificar o perfil genético do agressor (Peterson *et al.*, 2012). A identificação humana tem sido realizada pela análise tradicional de STRs multiplex autossômicos, mas também de cromossomo Y. Os marcadores de cromossomo Y representam marcadores de linhagem e são passados de geração a geração sem alterações, visto que >95% do cromossomo Y não sofre recombinação (Butler, 2012). Assim, exceto por mutações randômicas, pais, filhos e irmãos de uma mesma linhagem paterna compartilham o mesmo perfil haplotípico de cromossomo Y (Butler, 2012). Apesar da desvantagem de não-individualização, a análise de marcadores de Y permitem a amplificação específica de DNA masculino de amostras que apresentam mistura de DNA feminino e masculino, como o caso de amostras provenientes de agressão sexual onde se tem células espermáticas e vaginais misturadas (Sinha *et al.*, 2003). A genotipagem de material biológico deixado em locais de crime não serve apenas para condenar culpados, mas

também para libertar inocentes. Por meio da análise do banco de dados do *Innocence Project* - USA, verificou-se que de 194 condenados (sendo 91% condenados por estupro ou por estupro e homicídio) foram inocentados por evidências de DNA (Hampikian *et al.*, 2011).

Enfim, amostras provenientes de material biológico em estado de conservação ruim, em escassez, com misturas, ou antigos podem apresentar pouco DNA útil e/ou em estado elevado de degradação de forma que não seja suficiente para a análise tradicional por STRs. Se a amplificação deste DNA com marcadores comerciais padrões tem eficácia insuficiente, portanto, o uso de SNPs passa a ser uma ferramenta alternativa e importante para esse fim. Uma vantagem da análise de SNPs sobre a de STRs é a maior eficácia na amplificação de poucas cópias do DNA molde devido ao seu *amplicom* com tamanho reduzido (~100pb). Outras vantagens estão diretamente associadas à tecnologia de análise, especialmente se ela for realizada com sistemas de nova geração (sequenciamento de nova geração; NGS).

2.3 ANÁLISE DE SNPS NA GENÉTICA FORENSE

Devido ao menor custo, facilidade de estudo e baixas taxas de mutação, os polimorfismos de nucleotídeo único (SNPs) tornaram-se componentes importantes no campo da genética forense, despertando interesse no desenvolvimento de painéis para avaliação de DNA baseados em SNPs (Gil, 2001). A utilização da metodologia de análise de SNPs é uma ferramenta adicional na identificação humana, na análise de ancestralidade e de predição de fenótipo com alto poder discriminatório para cada conjunto específico de SNPs, e já tem importância no cenário científico figurando como uma metodologia inovadora na elucidação de casos forense cíveis ou criminais (Phillips *et al.*, 2007; Tomas *et al.*, 2008).

Os SNPs são considerados polimorfismos genéticos devido a uma alteração em um único nucleotídeo em uma determinada posição do genoma humano e, em geral, são bialélicos. Os polimorfismos genéticos do tipo SNP correspondem a formas alternativas de sequências de nucleotídeos e, quando analisados no nível de DNA, diferem entre si por alterações do tipo substituição de base. Mais de quinze milhões de SNPs já foram descritos no genoma humano, e sua identificação pode ser feita pela ligação a segmentos flanqueantes à posição de um SNP específico, selecionando

apenas um curto fragmento para análise (Sobrinho *et al.*, 2005). Devido a sua natureza bialélica (em geral), um SNP possui um menor poder de discriminação individual se comparado com um STR multialélico. Assim, são necessários de quatro a oito SNPs para se ter uma equivalência ao poder de discriminação de um único *locus* STR (Krawczak, 1999). Estima-se que seja necessário entre quarenta e cinquenta SNPs para se alcançar o mesmo poder de discriminação individual que é fornecido pelo conjunto de dezesseis STRs, já em uso pela comunidade científica, para fins de individualização humana (Kidd *et al.*, 2006).

SNPs, que representam a classe mais abundante de variação genômica em humanos, podem servir tanto para a discriminação individual como para a predição das CEV. Sua abundância é a principal razão do grande interesse atual no campo forense; e o mapeamento do genoma humano tornou possível o desenvolvimento de mapas de genótipos e de haplótipos que podem caracterizar fenotipicamente um ser humano. Variações do tipo SNPs, podem levar a substituições de aminoácido na proteína, alterando as propriedades funcionais do produto traduzido, o qual pode se expressar em fenótipos distintos (O'rahilly, 2009). Entre as CEV, existem alguns estudos envolvendo a predição de características/traços faciais (Claes *et al.*, 2014), porém, os traços fenotípicos mais promissores e melhor estudados até agora para a identificação forense são aqueles relacionados à pigmentação da pele, olhos e cabelos, por serem características muito marcantes, de fácil visualização, e por constituírem um dos fenótipos mais variáveis na população humana (Kayser e Schneider, 2009). O desenvolvimento de técnicas quantitativas para mensurar a pigmentação proporcionou uma maior objetividade na classificação da ampla variação de pigmentação existente, permitindo identificar de forma mais precisa os genes que de alguma maneira exercem influência sobre essa característica.

Apesar de ser promissora a utilização de marcadores genéticos de traços fenotípicos na identificação humana para fins forenses, a descrição de genes e do mecanismo pelo qual eles influenciam a definição das características fenotípicas é de difícil elucidação, uma vez que essas características apresentam um padrão de herança complexo, por serem determinadas por múltiplos genes (poligenia) e por sofrerem forte influência do meio ambiente (Sturm e Larsson, 2009). Não obstante, estudos demonstram que a base genética da variação normal da pigmentação é passível de ser decifrada se forem avaliadas as variantes genéticas adequadas (Liu *et al.*, 2009).

2.4 PREDIÇÃO DE FENÓTIPO DE PIGMENTAÇÃO HUMANA

Estudos de associação genótipo-fenótipo sobre a cor do cabelo, a cor da íris e a pigmentação da pele têm pesquisado variações de SNPs em genes diretamente envolvidos na síntese de pigmentos (Sturm, 2009; Walsh *et al.*, 2011a; Walsh *et al.*, 2011b; Walsh *et al.*, 2013). Todos eles sustentam que há uma necessidade grande por parte dos órgãos policiais de se possuir uma base de dados com padrões genéticos que possam auxiliar na solução de crimes, excluindo ou inserindo indivíduos em fase de investigações criminais que buscam identificar vítimas, envolvidos ou suspeitos. Especialmente no caso de não se ter informações sobre o perfil genético de referência (para confronto de alelos de STRs), a amostra questionada pode ser estudada para fins da predição do fenótipo do indivíduo (Kayser *et al.*, 2011).

2.4.1 Biologia da Pigmentação

Acredita-se que existam mais de 120 genes de alguma forma envolvidos nas vias de pigmentação, os quais agem em diferentes estágios do processo de produção da melanina (Branicki *et al.*, 2009). Tais genes agem em caráter quantitativo, influenciando cumulativamente nas diferentes tonalidades de coloração. Contudo, cerca de apenas poucas dezenas de genes têm alelos que definitivamente determinam os padrões extremos de pigmentação melânica, isto é, os padrões escuro e claro. A compreensão da biologia da pigmentação depende do entendimento dos eventos intramelanossomais. Os melanossomas (compartimentos sub-celulares, produzidos pelos melanócitos, que sintetizam e estocam polímeros de melanina) são os principais responsáveis pela coloração dos tecidos humanos por causa do seu conteúdo de melanina (Sulem *et al.*, 2007). Melaninas são produzidas em dois tipos distintos quimicamente: eumelanina (coloração castanho-preto) e feomelanina (coloração amarelo-avermelhada) (Ito *et al.*, 2006). Eumelaninas são escuras, opacas e altamente polimerizadas (formato oval), enquanto que as feomelaninas, por possuírem sulfidrila e cisteína em seus passos de conjugação, são mais leves e menos polimerizadas (formato esférico), o que ocasiona uma deposição desforme no lúmen. O acúmulo, a transferência, o conteúdo e a forma destes componentes estão diretamente ligados à intensidade da coloração de pele e de olho (Kondo e Hearing, 2011).

2.4.1.1 Pigmentação na Pele

A melanina é gradualmente produzida nos melanossomas que, quando maduros, são transportados ao longo das projeções dendríticas do melanócitos até os queratinócitos adjacentes da pele (Sturm, 2006). O número de melanócitos entre os diferentes tons de pele é constante, mas peles mais escuras apresentam melanossomas mais densos e individualmente dispersos, enquanto peles mais claras apresentam melanossomas menos densos e menores (Sturm *et al.*, 2012) (FIGURA 1). Os melanossomos desenvolvem gradualmente o pigmento enquanto amadurecem (Ho *et al.*, 2011). Após sintetizado, o pigmento é passado aos queratinócitos circundantes da pele e do folículo piloso (Sturm, 2006) e, por fim, para as células do epitélio pigmentar da retina (Kondo e Hearing, 2011).

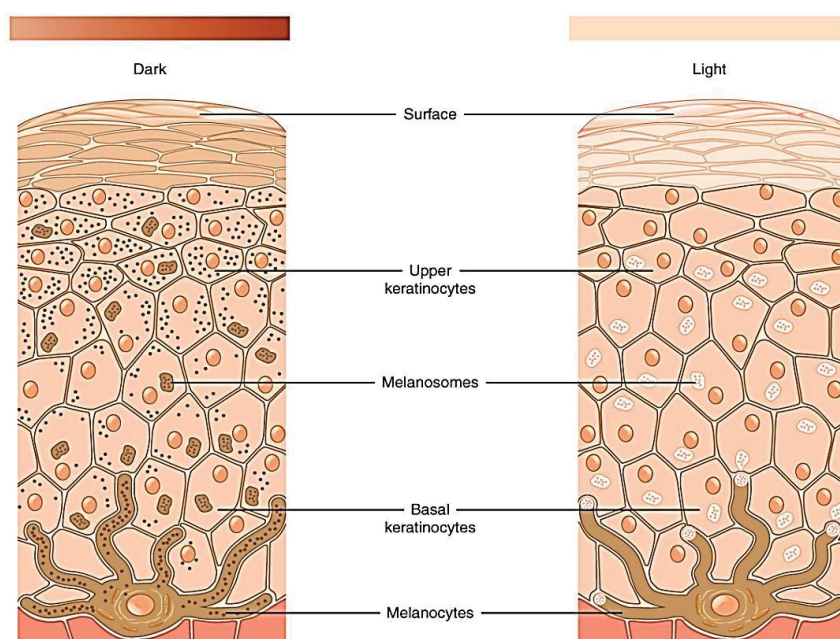


FIGURA 1: Produção e distribuição de melanina em peles escuras e claras

Fonte: <http://cnx.org/contents/471cb008-6900-476c-8a65-539fdc592b75@5>

O acúmulo de melanina e a distribuição do melanossomos variam quantitativamente entre indivíduos de diferentes ancestralidades (Sturm, 2006). Os melanossomos são divididos em quatro estágios de maturação (I a IV), determinados pela sua estrutura, quantidade, qualidade e arranjo da melanina produzida (Kushimoto *et al.*, 2001; Costin e Hearing, 2007) (FIGURA 2). O melanossomo no estágio I é uma

organela esférica, com ausência da enzima tirosinase (TYR) e de componentes internos estruturais. No estágio II a organela apresenta uma estrutura ovoide, atividade de TYR e a expressão de uma proteína da matriz estrutural, conhecida como PMEL17 (*premelanosome protein 17*), necessária para a produção da matriz fibrilar interna, e pela deposição mínima de melanina (Berson *et al.*, 2001; Costin e Hearing, 2007; Kondo e Hearing, 2011). No estágio III, a síntese de melanina é iniciada e o pigmento é depositado sobre fibras internas. No estágio IV de desenvolvimento, o melanossomo apresenta formato elíptico ou elipsoide. Estes estágios de desenvolvimento referem à eumelanossomos (que sintetizam a eumelanina), e nos feomelanossomos (que sintetizam a feomelanina) o processo é similar com a diferença que durante o processo de maturação, há ausência de fibrilas (Costin e Hearing, 2007; Brenner e Berking, 2010).

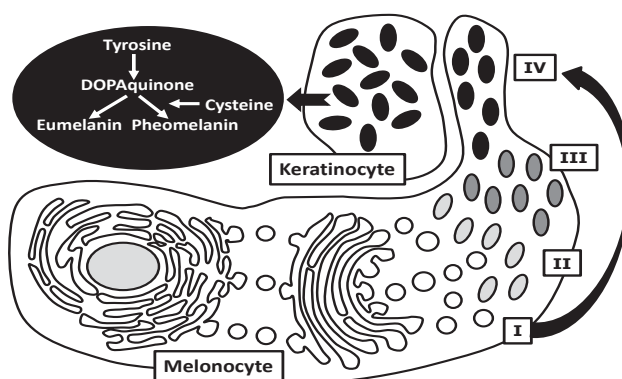


FIGURA 2: Esquema de um melanócito. O mecanismo para a produção da melanina. Os quatro estágios (I, II, III e IV) de desenvolvimento dos melanossomos são mostrados deslocando-se em direção à periferia da célula. À esquerda (ampliado) está um melanossomo maduro (estágio IV) onde se vê a via de síntese de eumelanina e feomelanina de forma simplificada.

2.4.1.2 Pigmentação no Olho

Os melanócitos dos olhos têm duas origens distintas, os que se encontram no epitélio pigmentoso da íris que têm origem neuroectodérmica, e os melanócitos estromais, que têm a mesma origem embrionária dos melanócitos dérmicos (oriundos da crista neural), e migram através do trato uveal durante o desenvolvimento. Nas íris de coloração castanha existe uma abundância de melanócitos e de melanina na camada basal anterior e no estroma, enquanto que nas íris de coloração azul essas camadas contêm muito pouco de melanina (Sturm e Larsson, 2009) (FIGURA 3).

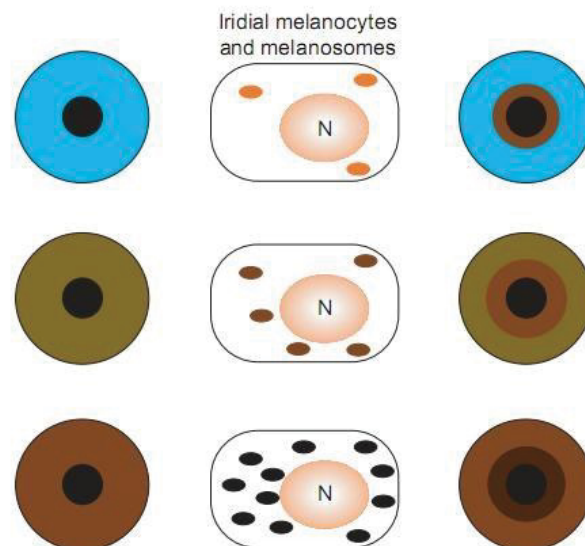


FIGURA 3: Pigmentação dos olhos. Diferenças entre os tipos e quantidades de melanossomos nos melanócitos dos olhos mais claros aos mais escuros.

Fonte: Sturm e Frudakis, 2004.

A íris, região ocular composta por tecido conectivo e muscular, é responsável pelo controle da entrada da luz incidente através de sua abertura central denominada pupila, regulando a formação das imagens na retina. Essa região é subdividida em cinco camadas, sendo a mais externa denominada borda anterior, seguida do estroma, esfíncter e músculos dilatadores e, mais internamente, o epitélio posterior pigmentado (Sturm e Frudakis, 2004). Dessas, a borda anterior e o estroma são as principais camadas contribuintes para a variação da coloração da íris (Wilkerson et

al., 1996), uma vez que o epitélio posterior pigmentado apresenta pigmentação em todas as cores observadas, não contribuindo de forma relevante para a variação fenotípica final (Sturm e Frudakis, 2004). Íris de coloração marrom apresentam melanócitos com alto conteúdo de melanina na borda anterior e no estroma, enquanto em íris azuis, essas camadas possuem baixo teor de melanina, permitindo maior passagem da luz e consequente refletância de ondas curtas azuis pelas fibras de colágeno presentes (FIGURA 4). No estroma, os dendritos dos melanócitos são geralmente paralelos à superfície da íris, os quais tendem a se agrupar na borda anterior, e correspondem a aproximadamente 66% das células componentes dessas camadas, independentemente da coloração da íris (Wilkerson *et al.*, 1996).

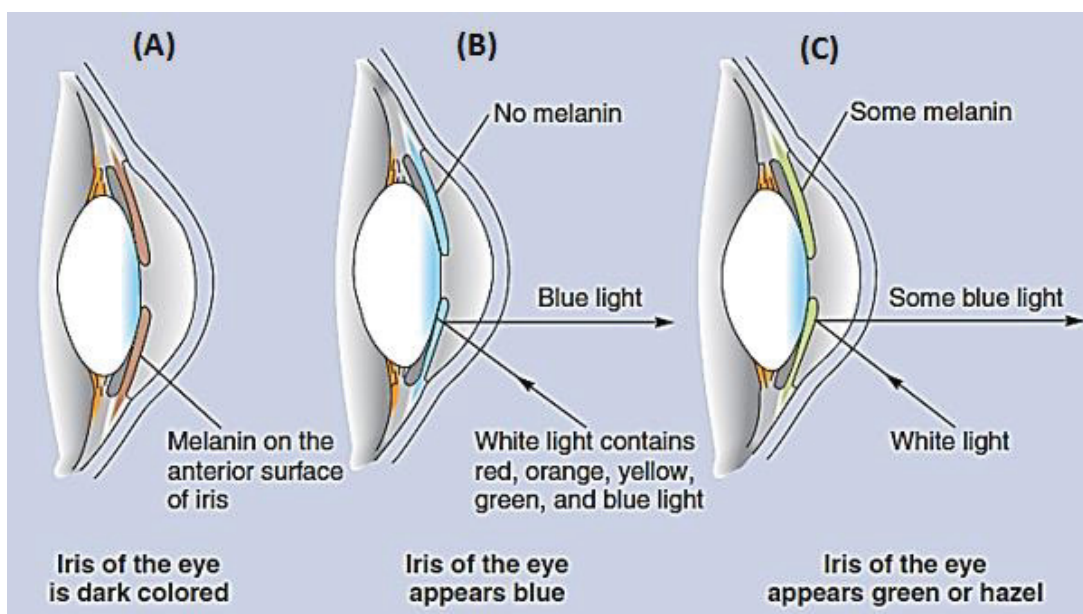


FIGURA 4: Epitélio pigmentoso da íris e refletância de ondas luminosas. (A) Olhos castanhos possuem bastante melanina na superfície da íris e absorvem todos os comprimentos de onda da luz. (B) Olhos azuis praticamente não possuem melanina na superfície da íris e refletem o comprimento de onda azul da luz. (C) Olhos verdes ou mel possuem uma quantidade intermediária de melanina e refletem parte do comprimento de onda azul da luz.

Fonte: <http://jurusanbiologi.blogspot.com.br/2014/04/the-inheritance-of-eye-color.html>

Ao contrário do que ocorre na pele, onde a melanina é continuamente produzida e secretada, os melanossomos da íris não são secretados dos melanócitos, mas sim retidos e acumulados no citoplasma destes. As variações na coloração observadas são resultado das diferentes quantidades e tipos de melanossomos presentes nos melanócitos, sendo a proporção de eumelanina/feomelanina geralmente maior em colorações mais escuras, e colorações mais claras apresentando maior concentração do pigmento feomelanina (Sturm e Larsson, 2009). Sabe-se que, apesar de todas as colorações de olhos possuírem um número similar de melanócitos, os olhos azuis contêm um mínimo de pigmentos e poucos melanossomos; os olhos verdes são resultado de moderados níveis de pigmentos, intensidade de melanina e número de melanossomos, enquanto olhos castanhos são resultado de níveis altos de melanina e grande quantidade de partículas melanossomais (Sturm e Frudakis, 2004) (FIGURA 5). Entretanto, estudos acerca das proporções eumelanina/feomelanina que possam vir a determinar a cor dos olhos são ainda inconclusivos.



FIGURA 5: Tonalidades de pigmentação de íris encontradas em humanos. Linha 1: Azul; Linha 2: Verde; Linha 3: Mel; Linha 4: Castanho.

Fonte: https://pt.wikipedia.org/wiki/Cor_dos_olhos#/media/File:Farbverlauf_Augenfarben.jp

2.4.1.3 Melanogênese

O regulador positivo da síntese de melanina mais importante é o receptor de melanocortina 1 (MC1R). Na sua via, o MC1R ativa a CREB (*cyclicAMP response-element binding protein*) que aumenta a expressão de MITF (*Microphthalmia-associated Transcription Factor*) e sua ativação, por fosforilação, estimula a transcrição de tirosinase (TYR), de proteína 1 relacionada a tirosinase (TYRP1) e de dopacromo tautomerase (DCT) (Liu *et al.*, 2013). O primeiro passo na melanogênese é a criação da dopaquinona (DQ) a partir da oxidação de tirosina pela tirosinase (TYR) com 3,4-dihidroxifenilalanina (DOPA) como agente intermediário (Kondo e Hearing, 2011) (FIGURA 6A).

Após a produção de DQ a sua espontânea ciclização acontece, dando origem a ciclodopa, que então faz uma reação de redução com uma molécula de DQ, resultando em uma molécula de DOPAcromo e uma de DOPA (Land *et al.*, 2003). O DOPAcromo, por sua vez, é decomposto espontaneamente por descarboxilação em pH neutro para formar 5,6-dihidroindole (DHI) e ácido 5,6-dihidroindole-2-carboxílico (DHICA) numa razão de 70:1 (Palumbo *et al.*, 1987). A enzima DOPAcromo tautomerase (DCT) tautomeriza o DOPAcroma produzindo somente DHICA (Palumbo *et al.*, 1991). DHI e DHICA são oxidadas e polimerizadas para dar origem a eumelanina (FIGURA 6B).

A feomelanogênese se dá pela adição de cisteína na oxidação da DQ para produzir dois isômeros principais de DOPAcisteinila (CD): a 5SCD e a 2SCD. Ocorre uma troca redox entre a CD e a DQ para produzir CD-quinonas e DOPA. A desidratação da CD-quinona faz com que ocorra a ciclização dando origem a orto-quinona (QI) que é rearranjada e forma o intermediário 1,4-benzotiazina, que se polimeriza dando origem a feomelaninas (Napolitano *et al.*, 1994; Napolitano *et al.*, 1999; Napolitano *et al.*, 2000; Greco *et al.*, 2009; Wakamatsu *et al.*, 2009) (FIGURA 6C).

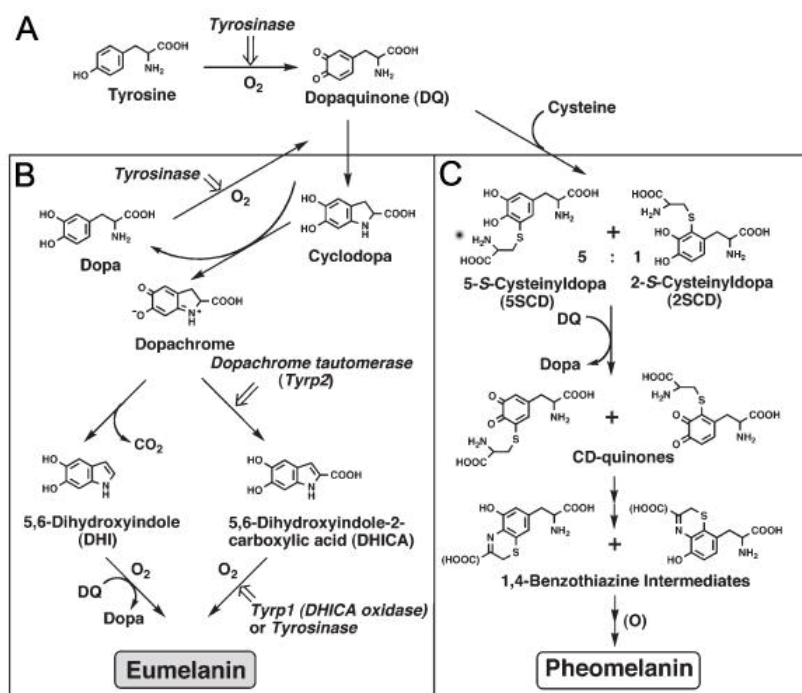


FIGURA 6 – Via biossintética da produção de eumelanina/feomelanina a partir de tirosina. As atividades da TYR, TYRP1 e DCT estão envolvidas na produção da eumelanina, mas apenas TYR e cisteína são necessários para a produção de feomelanina a partir de DQ. Fonte: Ito e Wakamatsu, 2010.

2.5 PIGMENTAÇÃO E ANCESTRALIDADE

Sabe-se que a pigmentação de pele está ligada à adaptação do indivíduo ao meio e por essa razão ser alvo de fatores evolutivos, como por exemplo, a seleção natural. Os seres humanos, diferente de outros primatas e mamíferos, não possuem pelos em todo o corpo. Uma explicação bem aceita à ausência de pelos é de que seria difícil a transpiração em lugares quentes e, assim, a regulação da temperatura seria mais eficiente na carência de pelos. Porém, a sobrevivência do indivíduo diante da quantidade reduzida de pelos na superfície do corpo implicaria na necessidade de uma proteção contra a radiação ultravioleta (UV). A principal via de proteção contra a luz UV é a presença de melanina na superfície da pele, a qual é capaz de absorver a radiação eletromagnética no comprimento das ondas ultravioleta resguardando,

assim, as moléculas de DNA, de proteínas e de outras macromoléculas (Rees, 2004; Liu *et al.*, 2013).

Observando a diferença da pigmentação e a localização geográfica das populações humanas, pode-se estabelecer uma relação entre radiação ultravioleta, perda dos pelos e alteração da pigmentação da pele: a pele negra oferece uma proteção maior em ambientes com alta incidência de radiação UV, como em locais de baixa latitude, enquanto a pele clara facilita a biossíntese de vitamina D em regiões de latitudes elevadas, onde a incidência de radiação UV é mediana ou baixa (Soejima *et al.*, 2007). Sendo assim, acredita-se que a história da transformação do fenótipo de pigmentação se deu, pelo menos em parte, pela pressão seletiva da radiação solar (Liu *et al.*, 2013).

Diante da relação fortemente positiva entre pigmentação e certos grupos populacionais, poder-se-ia questionar se a presença de determinado alelo seria apenas um marcador de ancestralidade genética ou se, de fato, poderia ele estar diretamente envolvido na determinação fenotípica da pigmentação. Na tentativa de solucionar esta questão, muitos genes têm sido estudados, e os produtos proteicos e/ou a expressão de suas variantes alélicas têm sido associados à síntese ou não de melaninas. Porém, no presente momento é possível identificar apenas alguns SNPs em genes os quais, de fato, estão diretamente envolvidos na síntese de pigmentos de pele e de olhos, independentemente da ancestralidade da população em estudo (Kenny *et al.*, 2012). Por exemplo, como será citado a seguir, o SNP rs12913832, presente no íntron 86 do gene *HERC2*, o qual se encontra a 21kb *upstream* do promotor de *OCA2*, regula a expressão de *OCA2* pelo dobramento da cromatina que permite o ancoramento dos fatores de transcrição de *OCA2*. O produto da expressão de *OCA2*, a proteína P, é essencial para a síntese de melanina. O alelo T (ou alelo A, considerando a fita complementar) do SNP rs12913832 ao permitir a abertura de cromatina, e o recrutamento dos fatores de transcrição de *OCA2*, induz à coloração mais escura, enquanto o alelo C (ou alelo G, considerando a fita complementar) mantém a cromatina mais fechada, tendo menos eficácia no recrutamento destes fatores, ocasionando uma coloração mais clara (FIGURA 7).

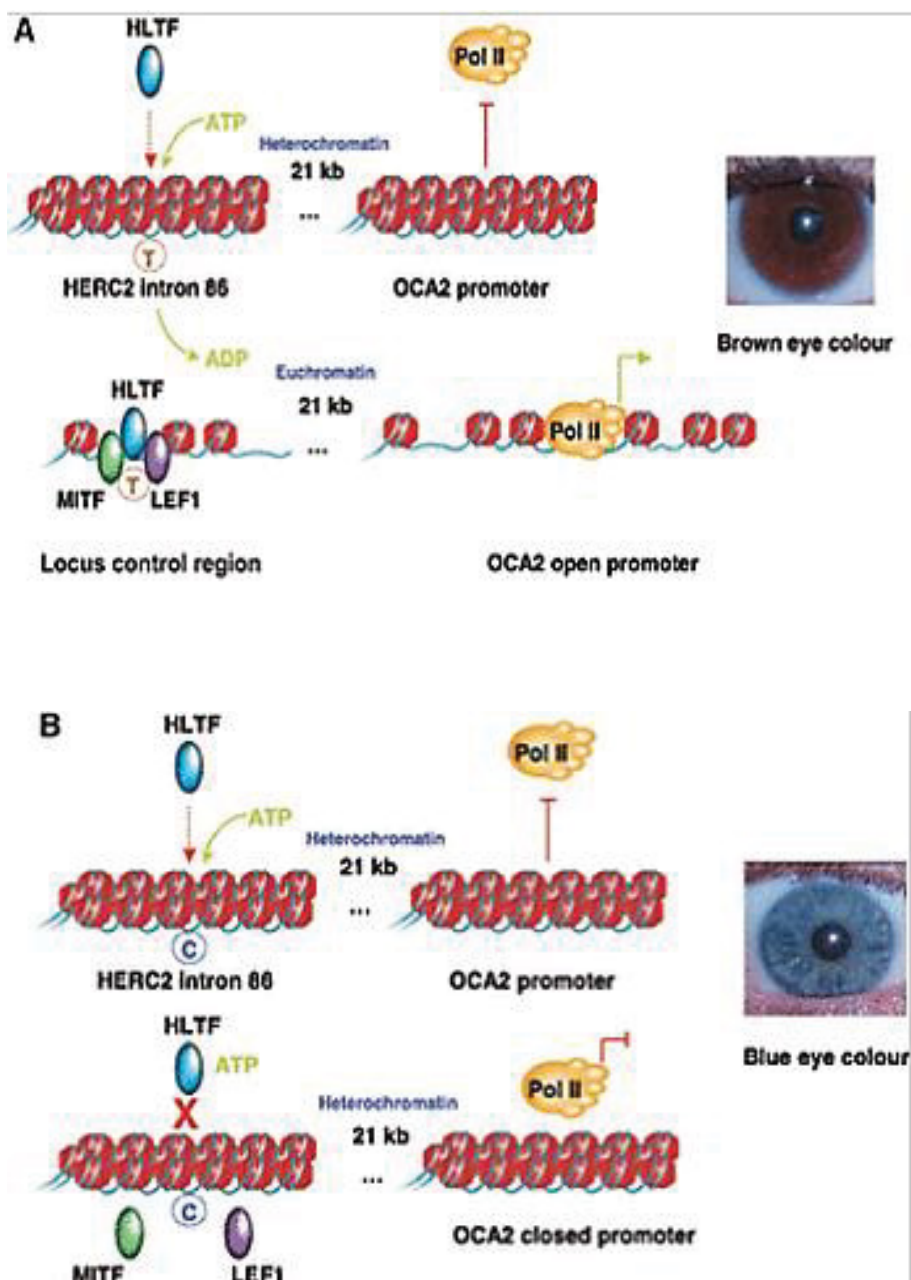


FIGURA 7: Modelo de determinação da cor de Iris castanho (A) e azul (B) influenciado pelo SNP rs12913832 do gene *HERC2*, segundo Sturm e Larsson, 2009.

Mesmo se localizando no interior de segmentos gênicos chave para a produção de melanina, poucos foram os alelos, contudo, já reconhecidos como sendo os próprios responsáveis pelo efeito variante de síntese de pigmento. Por não ter sido ainda completamente elucidada a implicação efetiva no fenótipo de cada variante alélica, algumas delas podem estar associadas à pele/olho escuro ou à pele/olho claro apenas por estarem em desequilíbrio de ligação com uma variante, de fato, fenótipo-efetiva. O desafio na análise da maioria dos SNPs ainda é saber se o marcador é de fenótipo ou de ancestralidade. Um exemplo é o caso do SNP rs26722 A>G do gene *SLC45A2* (regulador do pH que controla a via da síntese de melanina, no interior do melanócito). Algumas populações africanas de fenótipo escuro possuem o alelo A (Lys272) deste SNP, enquanto populações europeias de fenótipo claro possuem o alelo G (Glu272) com frequências similares às de aborígenes australianos de fenótipo escuro (frequência de G ~ 0,98) (Nakayama *et al.*, 2006). Devido ao fato do alelo G estar presente em grupos populacionais com diferentes fenótipos de pigmentação (europeus de pele clara e aborígenes de pele escura), e ao fato de populações africanas de cor escura conterem ambas variantes, Graf *et al.* (2007) concluíram que este polimorfismo não estaria associado com a determinação de pigmentação, mas sim que se trataria de uma variação associada à distribuição/isolamento geográfico. Assim, o alelo A surgiu como uma mutação recente apenas em africanos, e não interferiu na expressão da pigmentação da pele.

A população brasileira é miscigenada, com contribuições genéticas europeias, africanas e ameríndias, principalmente (Lins *et al.*, 2010). Isso se torna uma vantagem no estudo de SNPs relacionados à pigmentação, pois a existência de fenótipos muito diversos dentro de uma mesma população pode propiciar o encontro de SNPs funcionalmente efetivos no fenótipo e não relacionados à ancestralidade.

2.6 GENES E SNPS ENVOLVIDOS NA DETERMINAÇÃO DA PIGMENTAÇÃO

Mais de uma centena de genes com efeitos conhecidos já foram descritos em modelos animais e seus homólogos identificados em humanos (IFPCS, 2009). Esses genes podem ser divididos em grupos funcionais de acordo com a influência que seus produtos proteicos exercem na produção e regulação da pigmentação como segue:

- 1- genes que codificam fatores de crescimento e transcrição para o controle do

crescimento e diferenciação dos melanoblastos em melanócitos; 2- genes que codificam proteínas componentes dos melanossomos; 3- genes que controlam a biossíntese de organelas relacionadas; 4- genes que determinam a produção de eumelanina versus feomelanina e 5- genes envolvidos no transporte dos melanossomos.

Determinados SNPs nos genes *HERC2*, *OCA2*, *SLC24A5*, *SLC45A2*, *TYR*, *TYRP1* e *MC1R*, por exemplo, foram associados com a manifestação da pigmentação humana (Tully, 2007; Sulem *et al.*, 2007; Sturm *et al.*, 2008; Giardina *et al.*, 2011; Walsh *et al.*, 2011a; Walsh *et al.*, 2011b, Donnelly *et al.*, 2012; Allwood *et al.*, 2013; Liu *et al.*, 2013; Walsh *et al.*, 2013; Sawitzki *et al.*, 2017), sendo os três primeiros os responsáveis por grande parte da variação de coloração entre indivíduos de grupos populacionais em diferentes continentes (Liu *et al.*, 2013). Com base nesses estudos, oito SNPs foram selecionados para análise de predição de fenótipo de olhos e de pele (Sawitzki *et al.*, 2017), indicados no QUADRO 1. Na FIGURA 8, pode-se visualizar a localização na célula de cada produto gênico ou estrutura em questão.

QUADRO 1: Genes e SNPs candidatos à predição de pigmentação em humanos.

Gene	Lócus	Proteína	rs	Variação	SNP
<i>HERC2</i>	15q13.1	HERC2	rs916977	Íntron	A>G
			rs12913832	Íntron	A>G
<i>OCA2</i>	15q11.2	Proteína P	rs4778138	Íntron	G>A
<i>SLC24A5</i>	15q21.1	NCKX5	rs1426654	Exon	G>A
<i>SLC45A2</i>	5p13.2	MATP	rs16891982	Exon	C>G
<i>TYR</i>	11q14	TYR	rs1042602	Exon	C>A
<i>TYRP1</i>	9p23	TYRP1	rs2733832	Íntron	C>T
<i>MC1R</i>	16q24.3	MC1R	rs8045560	Íntron	C>T

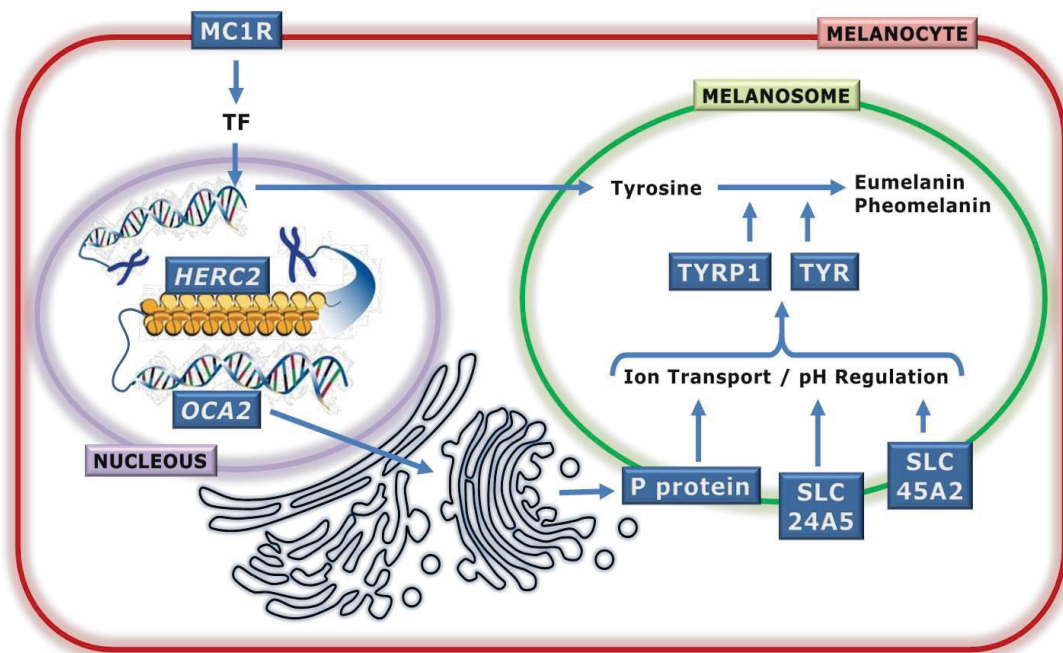


FIGURA 8: Indicação na estrutura celular dos genes e/ou seus produtos proteicos candidatos para predição de pigmentação em humanos, segundo Sawitzki *et al.*, 2017.

2.6.1 Gene *HERC2*

O gene *HERC2* (lôcus 15q13.1) codifica a proteína *HECT domain - RCC1-like domain containing protein 2* (ou *E3 ubiquitin-protein ligase HERC2*) envolvida no tráfego de proteínas. Ainda que a função do gene *HERC2* seja desconhecida, a sequência de DNA no interior de *HERC2* desempenha um papel estrutural no genoma e na regulação de pigmentação (Donnelly *et al.*, 2012). De acordo com Sturm *et al.* (2008), a proteína *HERC2* não está diretamente envolvida na síntese de pigmento, mas a sequência de DNA dentro da região do gene *HERC2* controla a expressão do gene *OCA2* o que, por sua vez, controla a produção de melanina (Sturm *et al.*, 2008). Tais sequências no interior de *HERC2* são *enhancers* do promotor do gene *OCA2* (Visser *et al.*, 2012). Visser *et al.* (2012) mostraram ainda que o SNP rs12913832 de *HERC2*, que se encontra a 21kb *upstream* do promotor de *OCA2*, regula a expressão deste gene pelo dobramento da cromatina que gera um *loop* de longo alcance que acaba por promover a expressão de *OCA2*. Esta atividade é mediada pelos fatores de

transcrição HLTF, LEF1 e MITF. Porém, diferentes versões alélicas de *HERC2* impedem o desenrolamento da cromatina, o que reduz (mas não impede) a ligação dos fatores de transcrição de *OCA2*, reduzindo consequentemente sua taxa transcricional. O alelo T (ou alelo A, considerando a fita complementar) do SNP rs12913832 permite a abertura da cromatina e consegue recrutar os fatores de transcrição de *OCA2*, o que permite a produção de melanina e ocasiona uma coloração mais escura, enquanto o alelo C (ou alelo G, considerando a fita complementar) apresenta uma baixa no recrutamento destes fatores, impedindo a síntese plena de pigmento e ocasionando uma coloração mais clara. Assim, via controle da expressão de *OCA2*, o gene *HERC2* tem o papel predominante na determinação de cor. Dados populacionais indicam que os SNPs rs916977 e rs12913832, localizados no íntron 12 e íntron 86, respectivamente, são melhores preditores de cor de olho do que três SNPs já descritos no interior do íntron 1 do próprio gene *OCA2* (rs7495174, rs6497268 e rs11855019) (Branicki *et al.*, 2009). Estudos em outras populações também indicam o SNP rs12913832 como sendo o SNP com maior efeito na predição da cor de olhos (Allwood *et al.*, 2013). Adicionalmente, se observou que o alelo A do SNP rs1667394, localizado no íntron 4 do gene *HERC2*, tem um padrão de associação com cor de cabelo e de olhos em um gradiente de redução da pigmentação: menor frequência alélica em indivíduos com cabelos e olhos castanhos e com a maior frequência em indivíduos com cabelo loiro e olhos azuis. O alelo A é encontrado em uma frequência de 80-90% na população do norte da Europa (Sulem *et al.*, 2007).

2.6.2 Gene *OCA2*

O gene que, ao conter mutações deletérias, ocasiona o albinismo oculocutâneo tipo 2 (gene *OCA2*), codifica uma proteína com 12 domínios transmembrana denominada proteína P (Sturm *et al.*, 2008). Acredita-se que a proteína P esteja envolvida no transporte de ânions, na regulação do pH melanossomal (Visser *et al.*, 2012), e que esteja envolvida no tráfego de proteínas internas como a tirosinase (TYR) e as proteínas associadas a tirosinase 1 (TYRP1) (Valenzuela *et al.*, 2011). *OCA2* por si só, independente de *HERC2*, também desempenha um papel na variação de coloração comum em olhos, cabelo e pele. Três SNPs que se encontram no intron 1 de *OCA2* mostraram ter uma forte associação com a coloração dos olhos (Liu *et al.*,

2013). Além disto, o SNP rs1800407 encontrado no exón 13 do gene *OCA2* ocasiona uma troca do tipo Arg419Gln na sequência polipeptídica da proteína P (Donnelly *et al.*, 2012), esta troca aumenta a penetrância do fenótipo de olhos azuis associados ao SNP rs12913832 (Sturm *et al.*, 2008; Walsh, *et al.*, 2011a). A proteína P, relacionada ao transporte de sódio/sulfeto (família SLC13), pode estar diretamente implicada na regulação do pH melanossomal (Kondo e Hearing, 2011). Foi observado que a proteína P atuava na permuta de cátions Na^+/H^+ , e/ou como um transportador de glutamato. Ambas as funções indicam que este produto do gene *OCA2* está envolvido no fornecimento de substrato para a TYR. A proteína P pode ainda estar envolvida no tráfego intracelular da enzima TYR, durante a maturação dos melanossomos (Eiberg *et al.*, 2008). O pH do melanossomo regula a atividade da TYR afetando, portanto, uma série de fatores na síntese da melanina, como a taxa de produção de melanina, a proporção de eumelanina/feomelanina e a maturação de melanossomos nos melanócitos e células de melanoma. Resultados experimentais indicaram que a produção de melanina é ideal em ambiente celular com pH próximo a pH-neutro (cerca de pH= 6.8). Paralelamente, a síntese da melanina é quase ausente em pH menor que 5.5. Em um pH menos ácido, por outro lado, a melanogênese pode prosseguir sem qualquer alteração na abundância de proteínas melanogênicas (Ancans *et al.*, 2001; Fuller *et al.*, 2001). Outro fator envolvido ainda na regulação do pH é a enzima ATPase-vacuolar (V-ATPase), presente no interior do melanossomo, a qual atua na acidificação dos melanossomos. Em estudo realizado por Fuller e colaboradores foi relatado que a V-ATPase está em menor quantidade ou, até mesmo, ausente em melanossomos em pele negra (Fuller *et al.*, 2001).

2.6.3 Gene *SLC24A5*

Baseado no que se conhece do gene *NCKX5* ($\text{Na}^+/\text{Ca}^{2+}/\text{K}^+$ exchanger 5), a regulação do pH melanossomal pode ser o papel da proteína codificada pelo gene *SLC24A5* (Tully, 2007). O gene *NCKX5* codifica uma proteína de 500 aminoácidos, que faz parte da família das proteínas envolvidas nas trocas de sódio/cálcio dependente de potássio, localiza-se próxima à membrana melanossomal e atua no transporte de membrana do melanossomo (Norton *et al.*, 2007; Liu *et al.*, 2013). Lamason *et al.* (2005) descobriram que *SLC24A5* está localizado em melanossomos ou seus precursores, e assumiram que ele poderia atuar acumulando Ca^{2+} no

melanossomo. O gene *SLC24A5* foi descoberto primeiramente em *zebra-fish*, no qual está relacionado ao fenótipo de uma listra clara significativa (Lamason *et al.*, 2005). O homólogo foi identificado e caracterizado em humanos (Lamason *et al.*, 2005). O gene humano mostrou que uma substituição de Ala111Thr, resultante do SNP rs1426654 (de Ala111-alelo G para Thr111-alelo A) tinha um efeito similar na pigmentação humana, isto é, conferia redução de melanina (fenótipo claro). De acordo com Ginger *et al.* (2008) o *SLC24A5* está associado com a rede trans-Golgi, eles também descobriram que o alelo associado com pele mais escura alelo G (Ala111) do SNP rs1426654 teve maior atividade de troca de íons em relação ao alelo A, o qual está associado com pele mais clara (Thr111). O alelo ancestral evolutivamente conservado (Ala111) do gene *SLC24A5* também foi encontrado em alta frequência em populações africanas e asiáticas, e o alelo derivado (Thr111) atingiu a fixação em europeus. Especificamente, a frequência para a variante Thr111 varia entre 99 e 100% entre as amostras da população de origem europeia, enquanto Ala111 tem uma frequência de 93 a 100% em amostras de africanos, nativos americanos e população do Leste Asiático (Dimisianos *et al.*, 2009).

2.6.4 Gene *SLC45A2*

Originalmente identificado como um antígeno do melanoma humano (AIM1) e recentemente renomeado carregador de soluto família 45, membro 2, o gene *SLC45A2* contém sete éxons que se estendem a uma região de aproximadamente 40kb e codifica a proteína MATP de 530 aminoácidos que supõe-se ter 12 domínios transmembrana (Graf *et al.*, 2007). Mutações no gene *SLC45A2* causam uma forma de albinismo em humanos (OCA4) (Vierkötter *et al.*, 2012; Tully, 2007; Sturm, 2006). Graf *et al.* (2005) relataram dois SNPs em *SLC45A2* que não estavam associados a doenças, mas sim associados com a variação normal da pigmentação humana, e também a variações entre diferentes populações: Glu272Lys (rs26722) e Phe374Leu (rs16891982). A proteína MATP desempenha um papel crucial no processamento e tráfego intracelular da tirosinase (TYR), uma das enzimas cruciais e necessárias para síntese de melanina, assim como na via de outras proteínas melanossomais (Branicki *et al.*, 2009; Vierkötter *et al.*, 2012). O SNP rs16891982 gera o polimorfismo proteico Phe374Leu e regula a função de transporte da região transmembrana da proteína MATP. A variante Leu374 (alelo G) desempenha um papel importante no transporte

de prótons, resultando em um pH ótimo intramelanossomal, o qual permite a atividade da tirosinase (TYR) e consequentemente a produção adequada da eumelanina (fenótipo pigmentado). A variante Phe374 (alelo C) pode alterar o transporte, o pH e a síntese do pigmento (Graf *et al.*, 2005; Tully, 2007; Vierkötter *et al.*, 2012). Nakayama *et al.* (2006) referiram o alelo C (Phe374) como alelo de fenótipo claro, já que ele resultou na redução da função da MATP, por alterar o tráfego intracelular de elementos melanossomais, criando um ambiente de decréscimo na produção de melanina. Estudando o mesmo gene, Graf *et al.* (2005) observaram que o alelo Lys272 do gene *SLC45A2* mostrou um significativo aumento nas populações asiáticas e afro-americanas quando comparadas a europeus. Contudo, as populações aborígenes australianas mostraram frequências alélicas similares às das populações europeias. Devido ao fato do fenótipo de pigmentação dos dois últimos grupos populacionais serem bem diferentes, os autores sugeriram que o polimorfismo rs26722 não estaria associado com a determinação de pigmentação, mas apenas seria uma variação associada à distribuição ou isolamento geográfico.

2.6.5 Gene *TYR*

O gene *TYR* transcreve a tirosinase, uma importante enzima para a fase inicial da melanogênese (Liu *et al.*, 2013). O gene *TYR* foi indicado como um loci para a susceptibilidade ao vitiligo generalizado (Jin *et al.*, 2012). Variações polimórficas em *TYR* estão diretamente associadas com alterações na coloração de olhos, cabelos e pele (Sulem *et al.*, 2007). Em especial, dois polimorfismos de *TYR*, rs1042602 (Ser192Tyr) e rs1126809 (Ala402Gly), têm mostrado frequências alélicas associadas a fenótipos de diferentes colorações: alelos que aparecem em alta frequência em populações europeias e estão ausentes em populações africanas (Stokowski *et al.*, 2007). As análises enzimáticas *in vitro* revelaram que existe uma redução de aproximadamente 40% na atividade catalítica de tirosinase devido à variação Ser192Tyr (Chaki *et al.*, 2011).

2.6.6 Gene *TYRP1*

Localizado no cromossomo 9p23, o gene *TYRP1* codifica a proteína tirosinase 1, que faz parte do complexo da enzima tirosinase durante a produção de melanina (Parra, 2007; Liu *et al.*, 2013). O impacto da proteína TYRP1, assim como a da DCT, na estabilidade da proteína TYR e na produção de outras enzimas envolvidas na formação catalítica de melanina mostra que alterações no gene *TYRP1* têm um papel importante na variação normal da pigmentação humana (Sturm, 2006). Há mais de dez anos, viram-se evidências que polimorfismos no *TYRP1* estão associados à coloração de olhos em europeus (Frudakis *et al.*, 2003). Estudos proteicos mostraram que a proteína codificada por *TYRP1* foi encontrada estando elevada 2,6 vezes na pele de africanos e de indianos (pigmentação escura) se comparada à pele de Mexicanos, Chineses e Europeus (pigmentação clara) (Alaluf *et al.*, 2003). Foi ainda observado que mutações raras (não polimórficas) em *TYRP1* são responsáveis pelo albinismo oculocutâneo tipo 3 (OCA3) (Sulem *et al.*, 2007).

2.6.7 Gene *MC1R*

Até o momento, o gene *MC1R* é o mais bem caracterizado dentre os genes descritos que influenciam a variação normal de pigmentação em humanos (Gerstenblith *et al.*, 2007). Este gene apresenta um padrão peculiar de distribuição dos seus alelos entre as populações humanas: populações africanas e outras caracterizadas pela pele escura apresentam menos polimorfismos desse gene, enquanto em populações asiáticas e europeias, esse gene é polimórfico (Makova e Norton, 2005). Estudos em diferentes populações têm demonstrado que a região codificadora do gene *MC1R* apresenta mais de 70 variantes já identificadas (Gerstenblith *et al.*, 2007), com impacto significativo no fenótipo de pigmentação nestes grupos populacionais (Harding *et al.*, 2000; Sturm, 2006). O gene *MC1R* transcreve para o receptor de melanocortina 1 (MC1R), uma proteína-G transmembrana localizada na membrana de melanócitos (Valenzuela *et al.*, 2011) e diretamente responsável pela regulação da síntese de eumelanina/feomelanina (Makova *et al.*, 2005). A ligação do hormônio paracrino alfa-estimulante dos melanócitos (α -MSH) ao gene *MC1R* causa uma cascata de sinalização de cAMP aumentando a produção de eumelanina (Makova *et al.*, 2005; Valenzuela *et al.*, 2011),

enquanto que a ligação de seu antagonista, a proteína de sinalização agouti (ASIP), resulta em menos produção de eumelanina e no aumento da produção de feomelanina (Liu *et al.*, 2013).

2.6.8 Outros Genes e SNPs relacionados à Pigmentação Humana

Com base na literatura disponível, foram selecionados 68 SNPs dentro de 21 genes (*ASIP*, *BNC2*, *DDB1*, *EXOC2*, *HERC2*, *IRF4*, *KITLG*, *LYST*, *MC1R*, *MFSD12*, *MYO5A*, *NPLOC4*, *OCA2*, *SLC24A4*, *SLC24A5*, *SLC45A2*, *TMEM138*, *TTC3*, *TYR*, *TYRP1* e *UGT1A6*) candidatos ao estudo de previsão de pigmentação em diversas populações, por estarem relacionados com a manifestação da pigmentação em seres humanos. No QUADRO 2, estão apresentadas informações acerca desses genes e SNPs.

QUADRO 2: Outros genes e SNPs candidatos à predição de pigmentação em humanos.

GENE		SNP	C R O M	POSIÇÃO	ALELO		REGIÃO	Freq. Europa*		Freq. Africa*	
					A1	A2		A1	A2	A1	A2
1	ASIP	rs2378249	20	34630286	A	G	INTRON	0.883	0.117	0.843	0.157
2	ASIP	rs2424984	20	34262569	C	T	INTRON	0.103	0.897	0.801	0.199
3	ASIP	rs6119471	20	34197406	G	C	INTRON	0.001	0.999	0.686	0.314
4	BNC2	rs10756819	9	16858086	G	A	INTRON	0.324	0.676	0.919	0.081
5	DDB1	rs11230664	11	61308900	C	T	INTRON	0.016	0.984	0.688	0.312
6	EXOC2	rs4959270	6	457748	C	A	INTERGÊNICA	0.524	0.476	0.799	0.201
7	HERC2	rs12913832	15	28120472	A	G	INTRON	0.364	0.636	0.972	0.028
8	HERC2	rs8039195	15	28270938	C	T	INTRON	0.214	0.786	0.797	0.203
9	HERC2	rs1129038	15	28111713	C	T	3' UTR	0.365	0.635	0.972	0.028
10	HERC2	rs3935591	15	28128866	T	C	INTRON	0.212	0.788	0.777	0.223
11	HERC2	rs916977	15	28268218	T	C	INTRON	0.234	0.766	0.899	0.101

12	HERC2	rs11636232	15	28141480	C	T	SINÔNIMA	0.678	0.322	0.985	0.015
13	HERC2	rs16950987	15	28281082	G	A	INTRON	0.910	0.090	0.343	0.657
14	HERC2	rs7494942	15	28118913	A	G	INTRON	0.244	0.756	0.890	0.110
15	HERC2	rs7170852	15	28182840	T	A	INTRON	0.220	0.780	0.794	0.206
16	HERC2	rs2238289	15	28208069	A	G	INTRON	0.801	0.199	0.298	0.702
17	HERC2	rs2240203	15	28249056	T	C	INTRON	0.910	0.090	0.349	0.651
18	HERC2	rs4932620	15	28269135	T	C	INTRON	0.016	0.984	0.085	0.915
19	HERC2	rs6497271	15	28120285	A	G	INTRON	0.016	0.984	0.626	0.374
20	IRF4	rs12203592	6	396321	C	T	INTRON	0.884	0.116	0.992	0.008
21	KITLG	rs10777129	12	88567936	A	G	INTRON	0.081	0.919	0.763	0.237
22	KITLG	rs642742	12	88905969	T	C	INTERGÊNICA	0.200	0.800	0.899	0.101
23	KITLG	rs12821256	12	88934558	T	C	REGULATÓRIA	0.879	0.121	0.998	0.002
24	LYST	rs3768056	1	235744525	A	G	INTRON	0.772	0.228	0.914	0.086
25	MC1R	rs3212345	16	89915864	C	T	INTRON	0.424	0.576	0.942	0.058
26	MC1R	N29insA	16	89985753	DEL	InsA	EXON	-	-	-	-
27	MC1R	rs885479	16	89919746	G	A	EXON	0.930	0.070	0.993	0.007
28	MC1R	rs1805005	16	89919436	G	T	EXON	0.888	0.112	0.995	0.005
29	MC1R	rs1805006	16	89919510	C	A	EXON	0.990	0.010	0.999	0.001
30	MC1R	rs1805009	16	89920138	G	C	EXON	0.992	0.008	0.998	0.002
31	MC1R	rs1110400	16	89919722	T	C	EXON	0.992	0.008	1.000	0
32	MFSD12	rs10424065	19	3545024	C	T	INTRON	0.991	0.009	0.702	0.298
33	MFSD12	rs6510760	19	3565255	G	A	INTRON	0.938	0.062	0.164	0.836
34	MYO5A	rs1724630	15	52415803	G	C	INTRON	0.811	0.189	0.417	0.583
35	NPLOC4	rs9894429	17	81629785	C	T	SINÔNIMA	0.588	0.412	0.185	0.815
36	OCA2	rs1800407	15	25903913	C	T	EXON	0.924	0.076	0.998	0.002
37	OCA2	rs7170989	15	27955262	C	T	INTRON	0.750	0.250	0.323	0.677
38	OCA2	rs1448484	15	28038295	G	A	INTRON	0.008	0.992	0.803	0.197
39	OCA2	rs1037208	15	27986211	T	G	INTRON	0.817	0.183	0.410	0.590
40	OCA2	rs1375164	15	28046666	T	C	INTRON	0.237	0.763	0.961	0.039
41	OCA2	rs1597196	15	28049776	T	G	INTRON	0.199	0.801	0.762	0.238
42	OCA2	rs1800404	15	27990627	C	T	SINÔNIMA	0.214	0.786	0.873	0.127
43	OCA2	rs1900758	15	27984951	C	T	INTRON	0.335	0.665	0.893	0.107

44	OCA2	rs2036213	15	27839034	C	A	INTRON	0.583	0.417	0.762	0.238
45	OCA2	rs2594935	15	27939892	A	G	INTRON	0.331	0.669	0.866	0.134
46	OCA2	rs3794606	15	28023862	C	A	INTRON	0.240	0.760	0.958	0.042
47	OCA2	rs4778137	15	28082689	G	C	INTRON	0.311	0.689	0.663	0.337
48	OCA2	rs4778138	15	28090674	G	A	INTRON	0.169	0.831	0.718	0.282
49	OCA2	rs4778241	15	28093567	A	C	INTRON	0.242	0.758	0.576	0.424
50	OCA2	rs4778232	15	28036619	T	C	INTRON	0.240	0.760	0.837	0.163
51	OCA2	rs895829	15	28052911	C	T	INTRON	0.245	0.755	0.962	0.038
52	OCA2	rs895828	15	28052887	C	G	INTRON	0.098	0.902	0.571	0.429
53	SCL24A4	rs12896399	14	92307319	G	T	INTERGÊNICA	0.570	0.430	0.979	0.021
54	SCL24A4	rs2402130	14	92334859	G	A	INTRON	0.213	0.787	0.600	0.400
55	SCL24A5	rs1426654	15	48134287	G	A	EXON	0.003	0.997	0.926	0.074
56	SCL45A2	rs16891982	5	33951588	C	G	EXON	0.062	0.938	0.964	0.036
57	SCL45A2	rs28777	5	33958854	C	A	INTRON	0.044	0.956	0.805	0.195
58	SCL45A2	rs13289	5	33986304	C	G	3' UTR	0.401	0.599	0.742	0.258
59	SCL45A2	rs183671	5	33964105	T	G	INTRON	0.047	0.953	0.790	0.210
60	TMEM138	rs7948623	11	61369675	A	T	REGULATÓRIA	0.994	0.006	0.734	0.266
61	TTC3	rs2835630	21	37149542	G	A	INTRON	0.476	0.524	0.654	0.346
62	TYR	rs1042602	11	89178528	C	A	EXON	0.628	0.372	0.988	0.012
63	TYR	rs1393350	11	89277878	G	A	INTRON	0.756	0.244	0.992	0.008
64	TYRP1	rs683	9	12709305	A	C	3' UTR	0.627	0.373	0.112	0.888
65	TYRP1	rs13289810	9	12396731	A	G	INTERGÊNICA	0.668	0.332	0.750	0.250
66	TYRP1	rs1325127	9	12668328	C	T	INTERGÊNICA	0.346	0.654	0.731	0.269
67	TYRP1	rs2733832	9	12704725	C	T	INTRON	0.428	0.572	0.946	0.054
68	UGT1A6	rs2070959	2	233693545	A	G	EXON	0.690	0.310	0.761	0.239

A1: alelo ancestral; A2: alelo derivado; *Frequência alélica na população europeia;

#Frequência alélica na população africana. Fonte: www.ensembl.org/index.html.

Vê-se, portanto, a literatura indicar que indivíduos que herdam alelos de genes cujos produtos induzem à síntese de pigmentos apresentam pigmentos escuros em pele e olhos, e sujeitos que herdam outras versões alélicas destes mesmos genes têm estruturas pouco pigmentadas. Alguns SNPs, de fato, influenciam produtos gênicos que interferem na síntese de pigmentos. Outros são neutros, mas podem estar em desequilíbrio de ligação com os primeiros. Uma terceira classe de SNPs pode não ter nem efeito sobre nem ligação ao fenótipo de coloração, mas são importantes marcadores de ancestralidade. O presente trabalho pretendeu investigar a associação do fenótipo de cor de olhos e de pele com variantes alélicas de SNPs em genes que estão envolvidos com a síntese de melanina, em população miscigenada do sul do Brasil.

3 RESULTADOS E DISCUSSÃO

Os resultados e a discussão estão apresentados em formato de artigo científico.

ARTIGO 1 - STATISTIC MODELS TO PREDICT EYES AND SKIN COLORS IN SOUTHERN BRAZILIAN SUBJECTS

(Submetido ao periódico *Forensic Science International: Genetics*)

ARTIGO 2 - BR-EASE: BRAZILIAN EYE AND SKIN ESTIMATOR. A FORENSIC WEBTOOL FOR DNA PHENOTYPING.

(Submetido ao periódico *Forensic Science International: Genetics*)

STATISTIC MODELS TO PREDICT EYES AND SKIN COLORS IN SOUTHERN BRAZILIAN SUBJECTS

Carlos Eduardo Ibaldo Gonçalves^{1,2}, Fernanda Rosa Sawitzki¹, Alessandro Kahmann³, Michel Jose Anzanello³, Eduardo Filipe Avila Silva^{1,4,6}, Marcelo Malaghini⁵, Maria da Graça Bicalho², Clarice Sampaio Alho^{1,6}

¹Laboratório de Genética Humana e Molecular, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brazil.

²Laboratório de Imunogenética e Histocompatibilidade, Universidade Federal do Paraná, Curitiba, Brazil.

³Laboratório de Otimização de Produtos e Processos, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

⁴Setor Técnico-Científico, Superintendência Regional da Polícia Federal do Rio Grande do Sul, Porto Alegre, Brazil.

⁵Laboratório de Genética Molecular Forense, Polícia Científica do Paraná, Curitiba, Brazil.

⁶Instituto Nacional de Ciência e Tecnologia Forense, Porto Alegre, Brazil.

CORRESPONDENCE TO

Carlos Eduardo Ibaldo Gonçalves, Laboratório de Genética Humana e Molecular, Escola de Ciências, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Av. Ipiranga, 6681 – Prédio 12C, sala 233, 90619-900, Porto Alegre – RS, Brasil.
Email adress: carlosibaldo@gmail.com

ABSTRACT

Genetic variants related to externally visible characteristics are of great relevance for eyes and skin colors prediction in humans. In this study, we tested the ability of a set of seven SNPs in pigment-related genes to predict eyes and skin colors in an admixed South Brazilian population. We genotyped 478 Southern Brazilian subjects with different skin and eye colors by SNaPshot chemistry and capillary electrophoresis. Furthermore, we designed five models (A, B, C, D, E) for eyes

phenotype prediction, using multinomial logistic regression, and additional three models (F, G, H) for skin phenotype prediction. Proposed A model considered five phenotypic classes for eyes (Blue, Green, Hazel, Light and Dark Brown: AIC 628.73 / ACC 0.75). Proposed B model considered three phenotypic classes for eyes (Blue, Intermediate and Dark Brown: AIC 440.73 / ACC 0.82). Proposed C, D and E models considered two phenotypic classes for eyes (Blue and Non-Blue: AIC 194.67 / ACC 0.94; Brown and Non-Brown: AIC 190.95 / ACC 0.91; Light and Non-Light: AIC 146.78 / ACC 0.95, respectively). Proposed F model considered six phenotypic classes for skin (White, Pale, Beige, Light Brown, Medium Brown and Dark Brown: AIC 1049.39 / ACC 0.49). Proposed G model considered three phenotypic classes for skin (Light, Intermediate and Dark: AIC 429.33 / ACC 0.82). Finally, proposed H model considered two phenotypic classes for skin (Light and Dark: AIC 177.11 / ACC 0.92). These models were able to predict eye and skin pigmentation with high assertiveness in subjects presenting light or dark phenotypes, while intermediate phenotypes subjects showed lower classification efficiency levels. Increase in accuracy was achieved as we grouped the categories. It is important for law enforcement agencies to have phenotype predictors based on local data to assist in their investigations, mainly because Brazilian population shows high levels of admixed ancestry. We believe this forensic DNA phenotyping system can be particularly useful in instances where classic DNA profiling fails to generate a database match, or to help classifying degraded corpses, skeletal, or other biological evidences, as and aid for missing person identification cases.

KEYWORDS

Human Pigmentation, Eyes color, Skin color, SNP, SNaPshot, Forensic DNA Phenotyping

INTRODUCTION

Forensic DNA phenotyping (FDP) is a novel field within forensic genetics that aims to predict appearance traits of an individual based on DNA markers. Efficient FDP technologies would be especially useful in cases where there are no reference sample [1-3]. Externally visible characteristics (EVCs) are complex biological features, which makes it more difficult to identify genes involved in their manifestation [4]. Eyes and skin colors are highly heritable genetic traits and the most obvious and distinguishable EVCs to be used in human identification. Genotype-phenotype association studies have focused in detecting single nucleotide polymorphisms (SNPs) within genes directly or indirectly involved in pigments synthesis [1-7]. SNPs are changes (substitutions, deletions or insertions) into a single nucleotide position in a DNA sequence. These point modifications may occur in human genome within coding regions or in sequences with regulatory function. However, most of the time they are located in intergenic spaces, without determined function yet [8-10].

There are hundreds of genes involved in the pigmentation process; for forensic purposes, however, a small number of SNPs have already been proved to be sufficient in the prediction of pigmentation patterns with high confidence [1,11]. In this work, seven SNPs, selected in previous publication, were analysed: *HERC2* (rs12913832), *OCA2* (rs4778138), *SLC24A5* (rs1426654), *SLC45A2* (rs16891982), *TYRP1* (rs2733832), *TYR* (rs1042602), and *MC1R* (rs8045560) [12]. These SNPs were previously shown to significantly alter melanin content in human melanocytes, supporting their functional role in pigmentation [13].

We investigated this set of SNPs and their efficiency to predict eyes and skin color, for forensic purposes, in individuals with different phenotypes from Southern Brazil, using multinomial logistic regression models.

MATERIALS AND METHODS

This project was approved by Pontifical Catholic University of Rio Grande do Sul Research Ethics Committee (CEP) under protocol #11-05722 (Of.CEP-0295/12 and Of.CEP-1041/12). An informed written consent and a survey form on sex, age,

self-declared biogeographical ancestry, place of birth and residency was completed and signed by all voluntary participants.

Oral swabs were collected from 478 individuals from Southern Brazilian population.

Digital photographs of subjects' eyes were taken and their colors were classified as: Blue (156), Green (36), Hazel (26), Light Brown (41), or Dark Brown (219). In addition, the amounts of red (R), green (G), and blue (B) color values were measured by analyzing each photo with COLORS software [14].

Each participant skin color was identified using the Fitzpatrick Score (Types 1 to 6) and classified as: White (137), Pale (161), Beige (56), Light Brown (41), Medium Brown (53), or Dark Brown (30). Also, amounts of RGB color values were measured in an inner and hairless portion of the right arm (below elbow) using a color analyzer equipment ACR-1023 (Instrutherm, Brazil) [15-16].

Genomic DNA from oral swabs was extracted using the NucleoSpin Tissue kit (Macherey-Nagel Inc.) following manufacturer instructions or by organic extraction method [17]. SNPs amplification was performed by multiplex PCR in a final volume of 25 μ L containing 2-10 ng gDNA, 0.2 μ M of each primer, and 1X Multiplex PCR Master Mix (Qiagen Inc). Thermal cycler conditions consisted of 95°C/5min, 30 cycles of 94°C/30s, 57°C/90s and 72°C/90s, and a final extension of 72°C/10min. PCR products were purified using USB ExoSAP-IT PCR Product Cleanup (Affymetrix) according to the manufacturer's protocol. For SNP analysis, SNaPshot Multiplex System (Applied Biosystems) was used with seven SNaPshot primers. Reactions were performed in a final volume of 10 μ L following the manufacturer's protocol. Multiplex single base extensions were carried out for 28 cycles of 96°C/10s, 50°C/5s and 60°C/30s. SNaPshot products were treated with FastAP (Thermo Scientific) according to the manufacturer's recommendation. 1 μ L of labeled products were mixed with 8.5 μ L of HiDi formamide and 0.5 μ L of Genescan-120 LIZ size standard (Thermo Scientific). They were then separated using an ABIPRISM 3130xl Genetic Analyzer (Applied Biosystems) with POP-4 matrix in a capillary 36cm. GeneMapper ID software version 4.0 (Applied Biosystems) was used to analyze allelic data (phenotyping and genotyping details are available in [12]).

Statistical analysis

In order to select variables (SNPs) that better categorizes genotypes according to their eyes or skin color, we runned an exhaustive search among genotyping data (i.e. we created a multinomial logistic model to any possible combination of available genetic elements) aiming to determine a variable subset that yields the better prediction model. Such evaluation was performed through the Akaike Information Criteria (AIC), an Information Criteria that analyses the model log-likelihood and penalizes addition of variables to find the most parsimonious model. Models presenting lower AIC values are determined as the better models [18].

According to Marquês *et al.*, prediction power of a specific model can be evaluated through a confusion matrix [19]. Considering a two classes response variable, all evaluation metrics can be derived from a 2x2 confusion matrix as shown in Table 1.

Table 1 – Confusion Matrix

Real	Predicted	
	Reference	Non-reference
Reference	True Positive (TP)	False Negative (FN)
Non-reference	False Positive (FP)	True Negative (TN)

From such matrix, sensitivity (true positive rate), specificity (true negative rate) and accuracy (correct classification rate) can be calculated according eqn. (E1), eqn. (E2) and eqn. (E3) respectively.

$$\text{Sensitivity} = \text{TP}/(\text{TP}+\text{FN}) \quad (\text{E1})$$

$$\text{Specificity} = \text{TN}/(\text{FP}+\text{TN}) \quad (\text{E2})$$

$$\text{Accuracy} = (\text{TP}+\text{TN})/(\text{TP}+\text{TN}+\text{FP}+\text{FN}) \quad (\text{E3})$$

Following Liu *et al.* [2], we performed multinomial logistic regression analysis and other additional statistic testes using IBM SPSS Statistics 17 [20] and the R 3.3.2 software [21].

RESULTS AND DISCUSSION

Eyes Color

In **Model A**, a comparison was performed among all five eyes color classes: Blue, Green, Hazel, Light Brown and Dark Brown. The set of variables that yielded the best result was SNPs rs12913832, rs16891982 and rs1426654, with an AIC of 628.73 and an accuracy of 0.7510. Table 2 presents the confusion matrix of the 10-fold cross validation and it can be noticed that Blue and Dark Brown eyes have high sensitivity (0.9423 and 0.9589 respectively). However, eyes of intermediate colors (Green, Hazel and Light Brown) are widely misclassified, which can be caused by the unbalanced number of samples in each class according to Maione *et al.* [22].

Table 2 – **Model A**: Eyes classification results with five classes

		Real				
		Blue	Green	Hazel	Light Brown	Dark Brown
Predicted	Blue	147	20	1	0	0
	Green	0	1	0	2	0
	Hazel	1	2	1	4	7
	Light Brown	1	1	2	0	2
	Dark Brown	7	12	22	35	210
Total Accuracy		0.7510				

In order to address such drawback, in **Model B** we merged Green, Hazel and Light Brown into a single “Intermediate” class and rerun the exhaustive search to find the most discriminatory SNPs for these new classes. Resulting in AIC value of 440.73, a set comprised of SNPs rs4778138, rs12913832, rs1426654 and rs2733832 was selected. Table 3 depicts the confusion matrix of the 10-fold cross validation after merge. Grouping some phenotypic classes improved total accuracy (0.8201), especially through increase of Intermediate class sensitivity (0.4951). Number of Intermediate samples misclassified as Blue does not have significant change when compared to the 5-classes separation (21 to 22); however, Intermediate-colored eyes classified as Dark Brown is drastically smaller (69 to 30). Regarding other eye colors,

despite Blue class maintain its sensitivity (0.9423), Dark Brown class samples are most commonly misclassified as Intermediate (9 to 25; 0.8858).

Table 3 –**Model B**: Eyes classification results with three classes

		Real		
		Blue	Intermediate	Dark Brown
Predicted	Blue	147	22	0
	Intermediate	7	51	25
	Dark Brown	2	30	194
Total Accuracy		0.8201		

To verify, in **Model C**, if Blue colored eyes have better discrimination power than other colors, we performed a comparison considering two classes using a reference class BLUE against the other class NON-BLUE. In this model, three SNPs were selected: rs4778138, rs12913832 and rs16891982, yielding an AIC of 194.67 and an accuracy of 0.9393.

To verify, in **Model D**, if Brown (Light and Dark Brown) colored eyes can be more effectively distinguished from other colors (Blue, Green and Hazel), we performed a comparison considering two classes using a reference BROWN class against alternative NON-BROWN class. As for the usage of Brown as reference, rs4778138, rs12913832, rs16891982, rs1426654 and rs2733832 SNPs were selected resulting into an AIC of 190.95 and an accuracy of 0.9121.

Considering phenotypic classification of Blue and Green classes can sometimes be difficult, due to borderline features where a precise class assignment is sometimes unclear, we design **Model E**, to test if Light (Blue and Green) colored eyes present a distinguish genetic pattern from other colors (Hazel, Light and Dark Brown). We performed a comparison considering two classes using a reference class LIGHT against the other class NON-LIGHT. In this model, three SNPs were selected: rs12913832, rs1426654, rs1042602, yielding and AIC of 146.78 and an accuracy of 0.9476.

Table 4 depicts the confusion matrix of the 10-fold cross validation among such group separations of Models C, D and E, and it can be noticed that two phenotype classes yielded better general accuracy than five or three classes. Using Blue colored

eyes as reference did not affect the number of Blue colored eyes misclassification. In Table 1 it can be noticed that only 13 Brown eyes samples are not classified in Brown classes; however, Table 3 indicates that the number of Brown eyes misclassified samples increase to 16 with the merge. Also, it can be noticed that using Light colored eyes as reference resulted in better accuracy, just one NON-LIGHT was erroneously classified and the problem of most green eyes being erroneously classified as blue declines.

Table 4 – **Models C, D and E**: Eyes classification results with two classes with different references

Model C - Blue referenced		Real	
		Blue	Non-Blue
Predicted	Blue	147	20
	Not Blue	9	302
Total Accuracy		0.9393	
Model D - Brown referenced		Real	
		Non-Brown	Brown
Predicted	Not Brown	192	16
	Brown	26	244
Total Accuracy		0.9121	
Model E - Light referenced		Real	
		Light	Non-Light
Predicted	Light	168	1
	Non-Light	24	285
Total Accuracy		0,9476	

Next, multinomial logistic regressions were conducted and these five models were tested to create eyes phenotype predictors. Based on SNPs rs12913832, rs16891982 and rs1426654 genotypes, we designed a phenotypic predictor to calculate the probability of hypothetical subjects to belonging to each eyes phenotype classes: Blue, Green, Hazel, Light Brown and Dark Brown (Model A). In a similar manner, based on rs4778138, rs12913832, rs1426654 and rs2733832 SNPs genotypes, we designed a second phenotypic predictor to calculate the probability of hypothetical subjects to be classified as one of the following eye color classes: Blue, Intermediate and Dark (Model B). Finally, three another phenotype predictors considering 2-classes categorization: Blue and Non-Blue (Model C, with SNPs

rs4778138, rs12913832 and rs16891982), Brown and Non-Brown (Model D, comprising rs4778138, rs12913832, rs16891982, rs1426654 and rs2733832 SNPs), and Light and Non-Light (Model E, including rs12913832, rs1426654, rs1042602 SNPs). Figure 1 illustrate the probability of a hypothetical set of genotypes to belonging to subjects with different eyes colors. We called 'H' (High melanin content) the alleles strongly associated with darker phenotype (rs12913832=A; rs4778138=G; rs1426654=G; rs16891982=C; rs2733832=C), and allele 'L' (Low melanin content) the alleles strongly associated with lighter phenotype (rs12913832=G; rs4778138=A; rs1426654=A; rs16891982=G; rs2733832=T). As shown in Figure 1, homozygote subjects for 'H' allele will always have Dark eyes, while homozygote subjects for 'L' allele will never have Dark eyes. In addition, heterozygote subjects are most likely to have Intermediate or Dark colored eyes. All five predictors are able to discriminate eyes phenotypes, with high sensitivity in subjects with light or dark phenotype, although obtained classification efficiency is smaller in subjects presenting intermediate pigmentation levels, with increasing accuracy as proposed categories are grouped together. Eye color phenotype predictor can be downloaded at the following link <http://sbcbr.inf.ufrgs.br/carbm/static/forensic/EYE-COLOR-PREDICTOR.zip>.

Human iris varies from dark brown to lighter shades of blue and this difference is determined by the type, amount and distribution of melanin in iris outer layer [23]. Brown eyes have higher pigment levels, mainly eumelanin, and large amount of melanosomes; green eyes contain moderate melanin and melanosomes levels and blue eyes have low concentration of both melanin and melanosomes [23]. Blue and green eyes occurrence is practically limited to ethnical Europeans, while brown eyes are found in most of the human population throughout the world [24]. Several researches have produced significant advances in the genetic understanding of human eye color from genetic linkage analysis, candidate gene and GWAS studies, and identified SNPs in different genes associated with eye color variation in Europeans [24-28]. Brazil has an admixed population, with European, African and Amerindian ancestry [29], therefore justifying the need for phenotypic studies considering local populations. Genetic studies commonly use blue, green / intermediate and brown [3, 24] phenotypical classes, but we tested five different categorical groups.

HERC2 gene, located near *OCA2*, is the most important eye color-related gene and regulates *OCA2* expression [24, 27-28]. rs12913832 is considered the best polymorphism for eye color predictor, segregating blue eyes from other types, and

presenting GG recessive genotype as a blue phenotype marker [1, 3, 30, 31, 24, 27-28]. In our study, this SNP was present in all eye prediction models. Of all 156 blue-eyed individuals in this study, 147 displayed a GG homozygous genotype and nine showed GA heterozygous genotype in this SNP. All nine individuals erroneously classified as belonging to Intermediate or Dark phenotypes in the different categorizations show heterozygous GA profiles. All models considered rs12913832 as the single SNP with higher predictive power among all marker set. Since a few blue-eyed individuals were found to possess an A allele in rs12913832, which is the allele associated with eumelanin pigment production, our proposed predictor is classifying these individuals in darker phenotype classes. In addition, green-eyed individuals erroneously classified as blue also had the GG genotype in rs12913832. A thoroughly review of such misclassified subjects was performed, and we identify this samples as featuring particular phenotypes where category designation in blue or green were not obvious, due to borderline or ambiguous color tones and shades or eye features presented by this group of individuals. In such cases where an eye color classification is difficult or imprecise, even when detailed and clear images are available, additional categories would not provide useful information for forensic purposes. Therefore, we decided to group these two phenotypes into a LIGHT class, in model E, to reduce classification error and to better distinguish light eyes from more pigmented ones. rs4778138 was selected for B, C and D eye prediction models and it has been strongly associated with eye color in previous studies [5, 24, 27-28].

HERC2 / OCA2 region is the main determinant of eye color and provides more phenotypic information about blue and brown eyes [4, 24, 27-28]. However, new genomic association studies have been developed, aiming to identify alternative genes contributing on a smaller scale to eye phenotypes and, together, for determination of eye color. rs16891982 was selected for A, B, and D eye prediction models and is in agreement with Spichenok *et al* [31] and Liu *et al* [2]; the latter study served as basis for the first DNA forensic phenotyping system development, the Irisplex [3]. The Irisplex system achieves accuracy values of up to 0.93 for brown eyes and 0.91 for blue eyes in Dutch Europeans [3]. Our study showed values of accuracy of up to 0.95 for brown eyes and 0.94 for blue eyes, as well as similar issues dealing with classification of intermediate eye colors. Selection of rs1426654 (A, B, D and E models), rs1042602 (model E) and rs2733832 (B and D models) are in agreement with previous studies conducted in different populations [5, 31-33], that associated these SNPs with eye

color phenotype. rs8045560 was not relevant to discriminate eye phenotypes considering this sample set, maybe because *MC1R* is usually related to hair and skin phenotype [34, 35].

Despite the great advances in the genetic determination of eyes color phenotype, it is still a challenge to predict intermediate eyes colors. Thus, it is necessary to develop studies in order to find out if there are additional DNA variants, previously unknown or not associated with such physical trait, that specifically contribute to the expression of these characteristics, to increase accuracy of these methodologies and provide high reliability for forensic use.

Skin Color

For skin color analysis, in **Model F** we used six classes: White, Pale, Beige, Light Brown, Medium Brown and Dark Brown. Six variables (rs4778138, rs12913832, rs16891982, rs1426654, rs2733832 and rs1042602) have been found to contribute in the best generated model, which yields an AIC of 1049.39. Table 5 presents the confusion matrix for 6-classes skin color classification, and it can be noticed that such number of classes results in below 0.5 sensitivity in all non-white classes and a total accuracy of only 0.4874. Once again, questions regarding samples assignment in proposed categories might also be impacting overall accuracy, due to borderline skin tones phenotypical classification.

Table 5 – **Model F**: Skin classification results with six classes

		Real					
		White	Pale	Beige	Light Brown	Med Brown	Dark Brown
Predicted	White	89	78	3	1	2	0
	Pale	42	65	25	7	2	0
	Beige	4	12	21	5	1	0
	Light Brown	1	5	4	16	7	0
	Med Brown	1	1	2	11	27	15
	Dark Brown	0	0	1	1	14	15
Total Accuracy		0.4874					

To solve this question, as well as achieve better discrimination power, first we merge all six original classes into three: LIGHT (White and Pale), INTERMEDIATE (Beige and Light Brown) and DARK (Medium Brown and Dark Brown). With such

categorization in **Model G**, once again the best AIC (429.33) was reached employing the same six variables. Confusion matrix of the 10-fold cross validation, shown in Table 6, evidences improvements in total accuracy (0.8242) and a good sensitivity for light (0.9161) and dark (0.8554) skin groups, but INTERMEDIATE class still maintains high misclassification rates (0.5155 accuracy), suggesting that combination of Beige and Light Brown classes does not provide a well-defined class.

Table 6 – **Model G**: Skin classification results with three classes

		Real		
		Light	Intermediate	Dark
Predicted	Light	273	33	4
	Intermediate	23	50	8
	Dark	2	14	71
Total Accuracy		0.8242		

To better access genetic variations possibly involved in skin color pigmentation in Brazilian samples, in **Model H** we rerun the method with a simple 2-classes categorization: LIGHT (White, Pale and Beige) and DARK (Light, Medium and Dark Brown). With an AIC value of 177.11, once again the model with six variables was selected. As expected, the accuracy (0.9246) improved, but higher sensitivity levels were identified in the LIGHT class (0.9661), rather than in DARK class (0.8065), which comprehend a larger sample size (Table 7). Given the fact that all skin classifier models selected the same variables, it can be noticed that results considering six (Table 5), three (Table 6) and two classes (Table 7) are similar.

Table 7 – **Model H**: Skin classification results with two classes

		Real	
Dark referenced		Light	Dark
Predicted	Light	342	24
	Dark	12	100
Total Accuracy		0.9246	

Next, multinomial logistic regressions were conducted and these three models were tested in order to create skin phenotype predictors for Brazilian population. Based on SNPs rs4778138, rs12913832, rs16891982, rs1426654, rs2733832 and rs1042602 genotypes, we designed three phenotypic predictors to calculate the probability of hypothetical subjects to belonging to one of the following skin phenotypes: 1-White, Pale, Beige, Light Brown, Medium Brown and Dark Brown (Model F); 2-Light, Intermediate and Dark (Model G); and 3-Light and Dark (Model H). Figure 2 illustrate the probability of a hypothetical set of genotypes to be originated from subjects with different skin colors. We called 'H' (High melanin content) the alleles strongly associated with darker phenotype (rs12913832=A; rs4778138=G; rs1426654=G; rs16891982=C; rs2733832=C; rs1042602=C), and allele 'L' (Low melanin content) the alleles strongly associated with lighter phenotype (rs12913832=G; rs4778138=A; rs1426654=A; rs16891982=G; rs2733832=T; rs1042602=A). As shown in Figure 2, homozygote subjects for 'H' allele will always have Dark skin, while homozygote subjects for 'L' allele will never have Dark or Intermediate skin. In addition, heterozygote subjects are most likely to have Intermediate colored skin. Model F has a high AIC value, low accuracy, and does not discriminate well the six proposed categories; however it is possible to notice a clear segregation of Light and Dark phenotypes. Model G has lower AIC value and higher accuracy than Model F, and can separate better the three phenotypes, especially the extreme light and dark categories. However, skin Intermediate phenotypes prevails among mistakenly classified samples, indicating that Beige and Light Brown may not compose a homogeneous, well-defined class, and therefore cannot be grouped together. Genetically, Beige is closer to Light phenotype and Light Brown is closer to Dark phenotype. Model H presented the best results for skin phenotype prediction with a really low AIC value, when compared to the other models, and can segregate light skin from dark skin individuals with an estimated accuracy of 92.5%. Skin color phenotype predictor can be downloaded at the following link <http://sbc.b.inf.ufrgs.br/carbm/static/forensic/SKIN-COLOR-PREDICTOR.zip>.

Human skin pigmentation has a strong correlation with climatic distribution and ultraviolet radiation levels. Populations living closer to the Equator line generally have darker skin to protect themselves from ultraviolet radiation, which has higher incidence in this regions, and decrease folic acid loss, while populations at high latitudes have less pigmented skin to provide vitamin D production by absorbing the lowest ultraviolet

radiation in those regions [36-37]. Although skin color is related to climate adaptation, it is also known to be related to the melanogenesis process. Individuals with light or dark skin have similar numbers of melanocytes, but melanosomes are larger, more numerous and more pigmented in individuals with dark skin. Those with light skin possess less pigmented melanosomes, rich in light brown eumelanin and yellow / red pheomelanin [38].

The selection of these six variables by the multinomial logistic regression models was interesting given that genotype-phenotype associations found here were also observed in previous studies conducted in different populations. Liu *et al.* suggested that rs1426654 and rs16891982 are likely major predictors for skin color differences observed between continents [5] as well as Maroñas *et al.* [36]. Similar findings were also described by Spichenok *et al.* [31], and Mushailov *et al.* [39], which found association of rs1426654, rs16891982 and 12913832 with skin pigmentation. Likewise, Walsh *et al.* found association of rs1426654, rs16891982, rs12913832 and rs1042602 with skin pigmentation [40], and Chaitanya *et al.* used this study as a basis for formulating HIRISPLEX-S [41], which introduced skin phenotype prediction to HIRISPLEX DNA test system [35]. Selection of rs4778138 and rs2733832 are in agreement with previous studies conducted in different populations [5, 42-44], where these SNPs were associated with skin color phenotype. rs8045560, whose ancestral C allele (C>T SNP) is associated with darker skin populations [34], did not contribute to classes discrimination in our multinomial logistic regression models.

Available skin color phenotype prediction systems are the least refined to date, compared to hair and eyes color phenotype prediction counterparts, and it is still a challenge to predict skin intermediate colors [41]. Further characterization of polymorphisms associated with human skin pigmentation levels can lead to significant improvement in phenotype prediction, with major impact in forensic applications, and therefore must be pursued. Also, the relevance and influence of admixture levels found in specific populations must also be considered in phenotype prediction, since intermediate phenotypes consistently poses a challenge to precise category assignment in prediction models. In this context, local populations might also present discrepancies regarding genetic polymorphisms influencing or determining pigment levels, and forensic applications must be designed to address such differences.

CONCLUSION

Not rarely, global or foreign phenotype predictor systems fail in Brazilian samples evaluation due to high admixture levels presented by Brazilian population, in which genomic structure may favor epistatic and pleiotropic effects not observed in more homogeneous populations. Hence, it is important for law enforcement agencies to have access to phenotype predictors based on local data to assist in their investigations. The five eyes phenotype prediction models proposed are able to separate extreme phenotypes (Blue or Brown) with high assertiveness. In order to increase prediction accuracy and to reduce classification errors of intermediate phenotypes, we recommend the use of 2-classes model (E), which separates light eyes (Blue / Green) from pigmented ones. For skin color prediction through DNA, the 3-classes model (G) is shown to segregate individuals with Light, Intermediate or Dark skin, and the 2-classes model (H) was able to segregate individuals with Light or Dark skin, can be used. These DNA-based phenotype prediction models may be particularly useful in cases where classical DNA profile fails to generate database matching or to help classifying decomposed corpses, skeletons or other biological evidences and in cases of missing people identification.

FINANTIAL SUPPORT

This work was supported by CAPES Brazil (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Edital 25/2014 – Pró-Forenses), CNPq Brazil (Conselho Nacional de Pesquisa Científica - Chamada nº 16/2014 – INCT) and FAPERGS (Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul).

ACKNOWLEDGMENTS

The authors would like to thank Deborah Soares Bispo Santos Silva, Rodrigo Rodenbusch, Diego Wordell Gubert, Mayara Jorgens Prado, and Pietra Graebin for technical support.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] M. Kayser, Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes, *Forensic Sci. Int. Genet.* 18 (2015) 33-48. <http://doi.org/10.1016/j.fsigen.2015.02.003>.
- [2] F. Liu, K. van Duijn, J.R. Vingerling, A. Hofman, A.G. Uitterlinden, A.C.J.W. Janssens, M. Kayser, Eye color and the prediction of complex phenotypes from genotypes, *Curr. Biol.* 19 (2009) R192–R193. <http://doi.org/10.1016/j.cub.2009.01.027>.
- [3] S. Walsh, A. Lidenbergh, S.B. Zuniga, T. Sijen, P. Knijff, M. Kayser, K.N. Ballantyne, Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence, *Forensic Sci. Int. Genet.* 5 (2011) 464-471. <http://doi.org/10.1016/j.fsigen.2010.09.008>.
- [4] M. Kayser, P.M. Schneider, DNA-based prediction of human externally visible characteristics in forensics, motivations, scientific challenges, and ethical considerations, *Forensic Sci. Int. Genet.* 3 (2009) 154-161. <http://doi.org/10.1016/j.fsigen.2009.01.012>.
- [5] F. Liu, B. Wenb, M. Kayser, Colorful DNA polymorphisms in humans, *Semin. Cell. Dev. Biol.* 24 (2013) 562-575. <http://doi.org/10.1016/j.semcdb.2013.03.013>.
- [6] B. Budowle, A. Van Daal, Forensically relevant SNP classes, *Biotechniques* 44 (2008) 603-610. <http://doi.org/10.2144/000112806>.
- [7] E.A. Graham, DNA reviews: predicting phenotype, *Forensic Sci. Med. Pathol.* 4 (2008) 196-199. <http://doi.org/10.1007/s12024-008-9056-6>.

- [8] A.J. Brookes, The essence of SNPs, *Gene*. 234 (1999) 177-186. [https://doi.org/10.1016/S0378-1119\(99\)00219-X](https://doi.org/10.1016/S0378-1119(99)00219-X).
- [9] B. Budowle, SNP Typing Strategies, *Forensic Sci. Int.* 146 (2004) S139-S142. <https://doi.org/10.1016/j.forsciint.2004.09.042>.
- [10] R. Javed and Mukesh, Current research status, databases and application of single nucleotide polymorphism, *Pakistan Journal of Biological Sciences*. 13 (2010) 657-663. <https://doi.org/10.3923/pjbs.2010.657.663>.
- [11] T.N. Frudakis, *Molecular photofitting, predicting ancestry and phenotype using DNA*, Elsevier, 2008. <https://doi.org/10.1016/B978-0-12-088492-6.X5001-4>.
- [12] F.R. Sawitzki, R. Rodenbusch, D.W. Gubert, M.J. Prado, D.S.B.S. Silva, E.F.A. Silva, C.S. Alho, Analysis of Eight SNPs in Pigment-Related Genes in South Brazilian Subjects with High or Low Skin and Eye Melanin Content, *SM J. Forensic Res. Criminol.* 1 (2017) 1008.
- [13] A.L. Cook, W. Chen, A.E. Thurber, D.J. Smit, A.G. Smith, T.G. Bladen, D.L. Brown, D.L. Duffy, L. Pastorino, G. Bianchi-Scarra, J.H. Leonard, J.L. Stow, R.A. Sturm, Analysis of cultured human melanocytes based on polymorphisms within the SLC45A2/MATP, SLC24A5/NCKX5, and OCA2/P loci, *J. Invest. Dermatol.* 129 (2009) 392-405. <https://doi.org/10.1038/jid.2008.211>.
- [14] I. Otaka, K. Kumagai, Y. Inagaki, M. Shimoyama, K. Saegusa, T. Hara, Simple and inexpensive software designed for the evaluation of color, *Am. J. Ophthalmol.* 133 (2002) 140-142. [https://doi.org/10.1016/S0002-9394\(01\)01213-2](https://doi.org/10.1016/S0002-9394(01)01213-2).
- [15] F.C. Parra, R.C. Amado, J.R. Lambertucci, J. Rocha, C.M. Antunes, S.D. Pena, Color and genomic ancestry in Brazilians, *Proc. Natl. Acad. Sci. USA*. 100 (2003) 177-182. <https://doi.org/10.1073/pnas.0126614100>.

- [16] T.B. Fitzpatrick, The validity and practicality of sun-reactive skin types I through VI, *Arch. Dermatol.* 124 (1988) 869–871. <https://doi.org/10.1001/archderm.1988.01670060015008>.
- [17] D. Liu, *Handbook of Nucleic Acid Purification*. (2009).
- [18] B. Hernando, M.V. Ibáñez, J.A. Deserio-Cuesta, R. Soria-Navarro, I. Vilar-Sastre, C. Martinez-Cadenas, Genetic determinants of freckle occurrence in the Spanish population: Towards ephelides prediction from human DNA samples, *Forensic Sci. Int. Genet.* 33 (2018) 38–47. <https://doi.org/10.1016/j.fsigen.2017.11.013>.
- [19] A.I. Marqués, V. García, J.S. Sánchez, Two-level classifier ensembles for credit risk assessment, *Expert Systems with Applications*. 39 (2012) 10916–10922. <https://doi.org/10.1016/j.eswa.2012.03.033>.
- [20] W.N. Venables, B.D. Ripley, *Modern Applied Statistics with S*. Fourth Edition. Springer, New York, 2002. <https://doi.org/10.1007/978-0-387-21706-2>.
- [21] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [22] C. Maione, B. Lemos, A. Dobal, F. Barbosa, R. Melgaço, Classification of geographic origin of rice by data mining and inductively coupled plasma mass spectrometry, *Comp. Electron. Agric.* 121 (2016) 101–107. <https://doi.org/10.1016/j.compag.2015.11.009>.
- [23] R.A. Sturm, T.N. Frudakis, Eye colour: portals into pigmentation genes and ancestry, *Trends Genet.* 20 (2004) 327–332. <https://doi.org/10.1016/j.tig.2004.06.010>.
- [24] M. Kayser, F. Liu, A.C. Janssens, F. Rivadeneira, O. Lao, K. van Duijn, M. Vermeulen, P. Arp, M.M. Jhamai, W.F. van Ijcken, J.T. den Dunnen, S. Heath, D. Zelenika, D.D. Despriet, C.C. Klaver, J.R. Vingerling, P.T. de Jong, A. Hofman, Y.S. Aulchenko, A.G. Uitterlinden, B.A. Oostra, C.M. van Duijn, Three Genome-wide Association Studies and a Linkage Analysis Identify HERC2 as a Human Iris Color

Gene, Am. J. Hum. Genet. 82 (2008) 411-423.
<https://doi.org/10.1016/j.ajhg.2007.10.003>.

[25] P. Sulem, D.F. Gudbjartsson, S.N. Stacey, A. Helgason, T. Rafnar, K.P. Magnusson, A. Manolescu, A. Karason, A. Palsson, G. Thorleifsson, M. Jakobsdottir, S. Steinberg, S. Pálsson, F. Jonasson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K.R. Benediktsdottir, K.K. Aben, L.A. Kiemeny, J.H. Olafsson, J. Gulcher, A. Kong, U. Thorsteinsdottir, K. Stefansson, Genetic determinants of hair, eye and skin pigmentation in Europeans, *Nat. Genet.* 39 (2007) 1443-1452.
<https://doi.org/10.1038/ng.2007.13>.

[26] T. Frudakis, M. Thomas, Z. Gaskin, K. Venkateswarlu, K.S. Chandra, S. Ginjupalli, S. Gunturi, S. Natrajan, V.K. Ponnuswamy, K.N. Ponnuswamy, Sequences associated with human iris pigmentation, *Genetics* 165 (2003) 2071-2083.

[27] H. Eiberg, J. Troelsen, M. Nielsen, A. Mikkelsen, J. Mengel-From, K.W. Kjaer, L. Hansen, Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the *HERC2* gene inhibiting *OCA2* expression. *Hum. Genet.* 123 (2008) 177-187. <https://doi.org/10.1007/s00439-007-0460-x>.

[28] R.A. Sturm, D.L. Duffy, Z.Z. Zhao, F.P. Leite, M.S. Stark, N.K. Hayward, N.G. Martin, G.W. Montgomery, A Single SNP in an Evolutionary Conserved Region within Intron 86 of the *HERC2* Gene Determines Human Blue-Brown Eye Color. *Am. J. Hum. Genet.* 82 (2008) 424-431. <https://doi.org/10.1016/j.ajhg.2007.11.005>.

[29] T.C. Lins, R.G. Vieira, B.S. Abreu, D. Grattapaglia, R.W. Pereira, Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs, *Am. J. Hum. Biol.* 22 (2010) 187-92.
<https://doi.org/10.1002/ajhb.20976>.

[30] M. Visser, M. Kayser, R.J. Palstra, *HERC2* rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer

and the OCA2 promoter, *Genome Res.* 22 (2012) 446-455.
<https://doi.org/10.1101/gr.128652.111>.

[31] O. Spichenok, Z.M. Budimlija, A.A. Mitchell, A. Jenny, L. Kovacevic, D. Marjanovic, T. Caragine, M. Prinz, E. Wurmba, Prediction of eye and skin color in diverse populations using seven SNPs, *Forensic Sci. Int. Genet.* 5 (2011) 472-478.
<https://doi.org/10.1016/j.fsigen.2010.10.005>.

[32] A. Vierkötter, U. Krämer, D. Sugiri, A. Morita, A. Yamamoto, N. Kaneko, M. Matsui, J. Krutmann, Development of lentigines in German and Japanese women correlates with variants in the SLC45A2 gene, *J. Invest. Dermatol.* 132 (2012) 733-736.
<https://doi.org/10.1038/jid.2011.350>.

[33] M. Chaki, M. Sengupta, M. Mondal, A. Bhattacharya, S. Mallick, R. Bhadra, Indian Genome Variation Consortium, K. Ray, Molecular and functional studies of tyrosinase variants among Indian oculocutaneous albinism type 1 patients, *J. Invest. Dermatol.* 131 (2011) 260-262. <https://doi.org/10.1038/jid.2010.274>.

[34] M.R. Gerstenblith, A.M. Goldstein, M.C. Fargnoli, K. Peris, M.T. Landi, Comprehensive evaluation of allele frequency differences of MC1R variants across populations, *Hum. Mutat.* 28 (2007) 495-505. <https://doi.org/10.1002/humu.20476>.

[35] S. Walsh, F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, W. Branicki, M. Kayser, The HirisPlex system for simultaneous prediction of hair and eye colour from DNA, *Forensic Sci. Int. Genet.* 7 (2013) 98-115.
<https://doi.org/10.1016/j.fsigen.2012.07.005>.

[36] O. Maroñas, C. Phillips, J. Söchtig, A. Gomez-Tato, R. Cruz, J. Alvarez-Dios, M.C. de Cal, Y. Ruiz, M. Fondevila, Á. Carracedo, M.V. Lareu, Development of a forensic skin colour predictive test, *Forensic Sci. Int. Genet.* 13 (2014) 34-44.
<https://doi.org/10.1016/j.fsigen.2014.06.017>.

- [37] N.G. Jablonski, G. Chaplin, Human skin pigmentation as an adaptation to UV radiation, *Proc. Natl. Acad. Sci.* 107 (2010) 8962–8968. <https://doi.org/10.1073/pnas.0914628107>.
- [38] G.S. Barsh, What Controls Variation in Human Skin Color?, *PLoS Biol.* 1 (2003) e7. <https://doi.org/10.1371/journal.pbio.0000027>.
- [39] V. Mushailov, S.A. Rodriguez, Z.M. Budimlija, M. Prinz, E. Wurmbach, Assay Development and Validation of an 8-SNP Multiplex Test to Predict Eye and Skin Coloration, *J. Forensic Sci.* 60 (2015) 990–1000. <https://doi.org/10.1111/1556-4029.12758>.
- [40] S. Walsh, L. Chaitanya, K. Breslin, C. Muralidharan, A. Bronikowska, E. Pośpiech, J. Koller, L. Kovatsi, A. Wollstein, W. Branicki, F. Liu, M. Kayser, Global skin colour prediction from DNA, *Hum. Genet.* 136 (2017) 847–863. <https://doi.org/10.1007/s00439-017-1808-5>.
- [41] L. Chaitanya, K. Breslin, S. Zuñiga, L. Wirken, E. Pośpiech, M. Kukla-Bartoszek, T. Sijen, P. de Knijff, F. Liu, W. Branicki, M. Kayser, S. Walsh, The HlrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation, *Forensic Sci. Int. Genet.* 35 (2018) 123–135. <https://doi.org/10.1016/j.fsigen.2018.04.004>.
- [42] M.P. Donnelly, P. Paschou, E. Grigorenko, D. Gurwitz, C. Barta, R. Lu, O.V. Zhukova, J. Kim, M. Siniscalco, M. New, H. Li, S.L.B. Kajuna, V.G. Manolopoulos, W.C. Speed, A.J. Pakstis, J.R. Kidd, K.K. Kidd, A global view of the OCA2-HERC2 region and pigmentation, *Hum. Genet.* 131 (2012) 683–696. <https://doi.org/10.1007/s00439-011-1110-x>.
- [43] D.L. Duffy, Z.Z. Zhao, R.A. Sturm, N.K. Hayward, N.G. Martin, G.W. Montgomery, Multiple Pigmentation Gene Polymorphisms Account for a Substantial Proportion of Risk of Cutaneous Malignant Melanoma, *J. Invest. Dermatol.* 130 (2010) 520–528. <https://doi.org/10.1038/jid.2009.258>.

[44] R.P. Stokowski, P.V.K. Pant, T. Dadd, A. Fereday, D.A. Hinds, C. Jarman, W. Filsell, R.S. Ginger, M.R. Green, F.J. van der Ouderaa, D.R. Cox, A Genomewide Association Study of Skin Pigmentation in a South Asian Population, *Am. J. Hum. Genet.* 81 (2007) 1119–1132. <https://doi.org/10.1086/522235>.

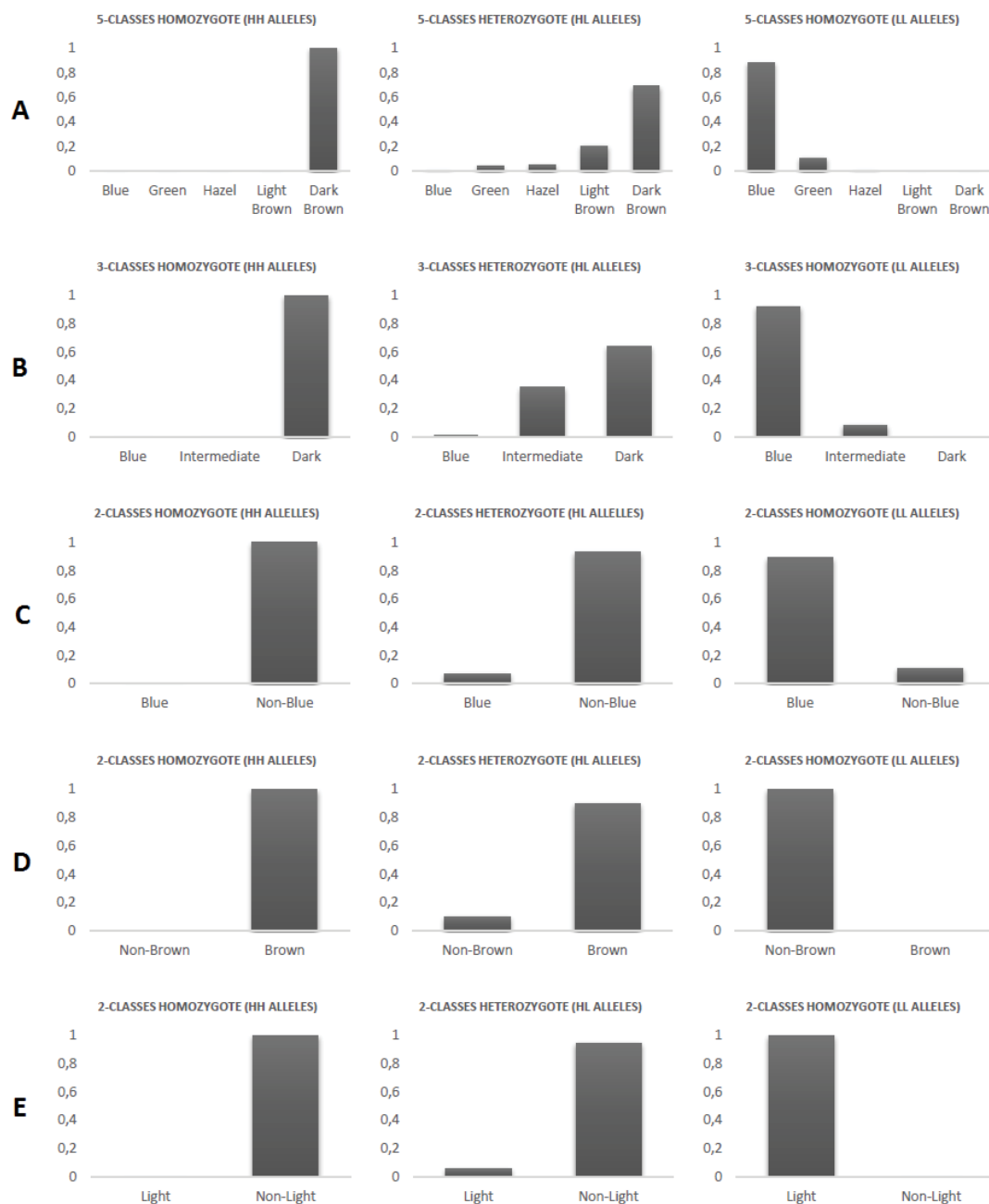


FIGURE 1 – Probability of a hypothetical set of genotypes to belong to subjects with different eyes colors in all five proposed models (A, B, C, D and E), considering all alleles in homozygosis (HH or LL) or in heterozygosis (HL). Allele ‘H’ (High melanin content) refers to alleles strongly associated with darker phenotype. Allele ‘L’ (Low melanin content) refers to alleles strongly associated with lighter phenotype.

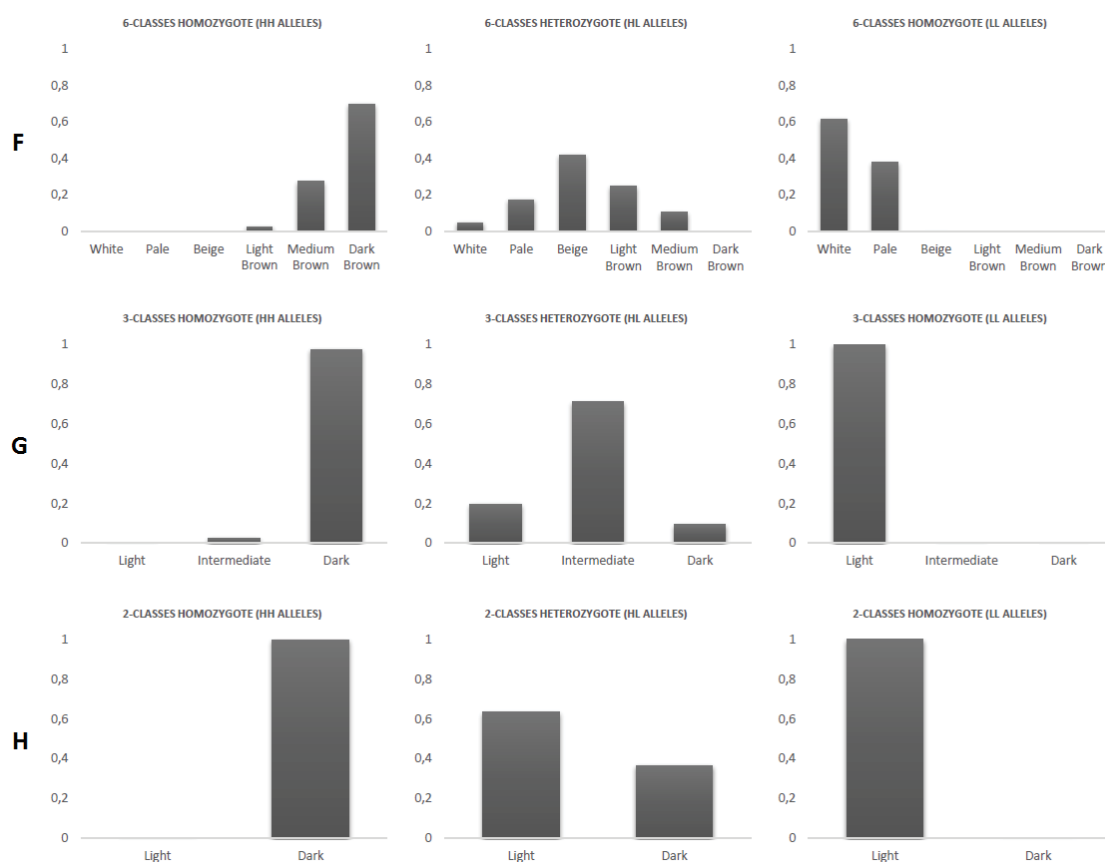


FIGURE 2 – Probability of a hypothetical set of genotypes to belong to subjects with different skin colors in all three proposed models (F, G and H), considering all alleles in homozygosis (HH or LL) or in heterozygosis (HL). Allele ‘H’ (High melanin content) refers to alleles strongly associated with darker phenotype. Allele ‘L’ (Low melanin content) refers to alleles strongly associated with lighter phenotype.

BR-EASE: BRAZILIAN EYE AND SKIN ESTIMATOR. A FORENSIC WEBTOOL FOR DNA PHENOTYPING

Carlos Eduardo Ibaldo Gonçalves^{1,2}, Márcio Dorn³, Eduardo Filipe Avila Silva^{2,4,6}, Marcelo Malaghini⁵, Clarice Sampaio Alho^{2,6}, Maria da Graça Bicalho¹

¹Laboratório de Imunogenética e Histocompatibilidade, Universidade Federal do Paraná, Curitiba, Brazil.

²Laboratório de Genética Humana e Molecular, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brazil.

³Structural Bioinformatics and Computational Biology Laboratory, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

⁴Setor Técnico-Científico, Superintendência Regional da Polícia Federal do Rio Grande do Sul, Porto Alegre, Brazil.

⁵Laboratório de Genética Molecular Forense, Polícia Científica do Paraná, Curitiba, Brazil.

⁶Instituto Nacional de Ciência e Tecnologia Forense, Porto Alegre, Brazil.

CORRESPONDENCE TO

Carlos Eduardo Ibaldo Gonçalves, Laboratório de Genética Humana e Molecular, Escola de Ciências, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Av. Ipiranga, 6681 – Prédio 12C, sala 233, 90619-900, Porto Alegre – RS, Brasil.
Email adress: carlosibaldo@gmail.com

ABSTRACT

Forensic DNA phenotyping (FDP) aims to predict appearance traits of an individual based on a DNA sample. Eyes and Skin colors are highly heritable genetic traits and are the most obvious and distinguishable externally visible characteristics to be used for human identification. Genotype-phenotype association studies have focused in detecting single nucleotide polymorphisms (SNPs) within genes directly or indirectly involved in pigments synthesis. We investigated 68 pigment production-related SNPs and their efficiency to predict eyes and skin color, for forensic purposes, in 465 admixed individuals with different phenotypes from Southern Brazil, using bioinformatics techniques. In this work, we introduce a simple system for eyes and skin phenotype prediction through DNA: BR-EASE, Brazilian Eye and Skin Estimator, available at '<http://sbcinf.ufrgs.br/brease>'. Eye color predictor uses 10 SNPs (*ASIP* rs6119471; *DDB1* rs11230664; *HERC2* rs1129038, rs12913832 and rs6497271; *MFSD12* rs10424065 and rs6510760; *OCA2* rs1800407; *SLC24A5* rs1426654; and *SLC24A4* rs2402130) to segregate three phenotype classes (blue, intermediate and brown) with an overall accuracy of 87%, and sensitivities of 98% for brown and 94% for blue, intermediate color has lower sensitivity of 31%. Skin color predictor uses 25 SNPs (*ASIP* rs2424984 and rs6119471; *DDB1* rs11230664; *HERC2* rs1129038, rs11636232, rs16950987, rs2238289, rs4932620 and rs6497271; *MC1R* rs1805006, rs1805009, rs3212345 and rs885479; *MFSD12* rs10424065 and rs6510760; *NPLOC4* rs9894429; *OCA2* rs1037208, rs1375164, rs3794606 and rs4778232; *SLC24A5* rs1426654; *SLC45A2* rs16891982 and rs183671; *TMEM138* rs7948623; and *TYR* rs1042602) to segregate three phenotypic classes (light, intermediate and dark) with a general accuracy of 97.5%, and sensitivities of 100% for light, 98% for dark and 90% for intermediate color. BR-EASE is a simple interface and easy to use tool that can be used by genetics and anthropological researchers, in the classroom as an exercise of multiple gene inheritance, and, especially, by law enforcement agencies to assist in their investigations.

KEYWORDS

BR-EASE, Human Pigmentation, Eyes color, Skin color, SNP, Forensic DNA Phenotyping

INTRODUCTION

Forensic DNA phenotyping (FDP) is a promising technique useful for filling gaps in unknown persons search, in criminal investigations, and in mass disaster victims or missing persons identification [1]. This methodology allows externally visible characteristics (EVC) prediction of an individual through the analysis of phenotype informative single nucleotide polymorphisms (SNPs) [2-3]. SNPs are changes (substitutions, deletions or insertions) into a single nucleotide position in a DNA sequence and these point mutations may occur in human genome within coding regions or in sequences with regulatory function. However, they are often located in intergenic spaces, without determined function yet [4-6].

Genotype-phenotype association studies have focused on human pigmentation features, since eye, hair and skin are the most obvious and distinguishable EVC to be used in human identification and are highly heritable genetic traits [1,7-21]. Also, these studies show that a relatively small number of genes is sufficient to provide most of phenotypic information of interest [1,22]. Several of these studies analyzed SNPs located in genes related to eye and hair pigmentation, obtaining high accuracy and sensitivity in the prediction of blue and brown eyes and red hair, while further research is needed to accurately predict other color phenotypes [13]. Skin prediction studies are more recent and have shown that a greater amount of SNPs is required for skin predicting [17], when compared to eye pigment-level inference.

In this work, we analyzed sixty-eight SNPs in twenty one genes reported in literature as associated with human pigmentation: *ASIP*, *BNC2*, *DDB1*, *EXOC2*, *HERC2*, *IRF4*, *KITLG*, *LYST*, *MC1R*, *MFSD12*, *MYO5A*, *NPLOC4*, *OCA2*, *SLC24A4*, *SLC24A5*, *SLC45A2*, *TMEM138*, *TTC3*, *TYR*, *TYRP1*, *UTG1A6* (see TABLE 1), in a population composed of 465 admixed individuals from Southern Brazil with different eye and skin colors.

Global or foreign phenotype predictors systems are usually based on European populations [2,16]. Brazil presents high admixing levels, with European, African and Amerindian ancestry [23], in which genomic structure may favor epistatic and pleiotropic effects not observed in ethnic homogenous groups. Therefore, it is important to analyze gene variants that specifically impact our population, in order to create predictors which can identify local individual phenotypic traits, aiming to assist law enforcement agencies in their investigations.

We investigated this set of 68 SNPs and their efficiency to predict eyes and skin color, for forensic purposes, in individuals with different phenotypes from Southern Brazil, using bioinformatics techniques. Eyes phenotypes were categorized into three classes: Blue, Intermediate and Brown. Skin phenotypes were also categorized into three classes: Light, Intermediate and Dark. Here, we introduce you to an easy online system for eyes and skin phenotype prediction through DNA: 'BR-EASE', Brazilian Eye and Skin Estimator. Available at '<http://sbcb.inf.ufrgs.br/brease>' server.

TABLE 1 - SNPs associated to human pigmentation related genes and its allelic frequencies in European, African and Southern Brazilian populations.

1	GENE	SNP	CROM	POSITION	ALLELES		VARIANT	FREQUENCY IN EUROPE *		FREQUENCY IN AFRICA *		FREQUENCY IN SOUTH BRAZIL		REFERENCES
					A1	A2		A1	A2	A1	A2	A1	A2	
1	ASIP	rs2378249	20	34630286	A	G	INTRONIC	0.883	0.117	0.843	0.157	0,885	0,115	13,17
2	ASIP	rs2424984	20	34262569	C	T	INTRONIC	0.103	0.897	0.801	0.199	0,229	0,771	1
3	ASIP	rs6119471	20	34197406	G	C	INTRONIC	0.001	0.999	0.686	0.314	0,112	0,888	1,17
4	BNC2	rs10756819	9	16858086	G	A	INTRONIC	0.324	0.676	0.919	0.081	0,422	0,578	7,17
5	DDB1	rs11230664	11	61308900	C	T	INTRONIC	0.016	0.984	0.688	0.312	0,145	0,855	18
6	EXOC2	rs4959270	6	457748	C	A	INTERGENIC	0.524	0.476	0.799	0.201	0,568	0,432	13,17
7	HERC2	rs12913832	15	28120472	A	G	INTRONIC	0.364	0.636	0.972	0.028	0,577	0,423	1,13,16,17
8	HERC2	rs8039195	15	28270938	C	T	INTRONIC	0.214	0.786	0.797	0.203	0,643	0,357	8
9	HERC2	rs1129038	15	28111713	C	T	3' UTR	0.365	0.635	0.972	0.028	0,572	0,428	1,17
10	HERC2	rs3935591	15	28128866	T	C	INTRONIC	0.212	0.788	0.777	0.223	0,369	0,631	1
11	HERC2	rs916977	15	28268218	T	C	INTRONIC	0.234	0.766	0.899	0.101	0,406	0,594	1
12	HERC2	rs11636232	15	28141480	C	T	SYNONYMOUS	0.678	0.322	0.985	0.015	0,778	0,222	1
13	HERC2	rs16950987	15	28281082	G	A	INTRONIC	0.910	0.090	0.343	0.657	0,787	0,213	8
14	HERC2	rs7494942	15	28118913	A	G	INTRONIC	0.244	0.756	0.890	0.110	0,421	0,579	9
15	HERC2	rs7170852	15	28182840	T	A	INTRONIC	0.220	0.780	0.794	0.206	0,376	0,624	10
16	HERC2	rs2238289	15	28208069	A	G	INTRONIC	0.801	0.199	0.298	0.702	0,789	0,211	10,17
17	HERC2	rs2240203	15	28249056	T	C	INTRONIC	0.910	0.090	0.349	0.651	0,799	0,201	10
18	HERC2	rs4932620	15	28269135	T	C	INTRONIC	0.016	0.984	0.085	0.915	0,025	0,975	18
19	HERC2	rs6497271	15	28120285	A	G	INTRONIC	0.016	0.984	0.626	0.374	0,128	0,872	18
20	IRF4	rs12203592	6	396321	C	T	INTRONIC	0.884	0.116	0.992	0.008	0,924	0,076	1,13,16,17
21	KITLG	rs10777129	12	88567936	A	G	INTRONIC	0.081	0.919	0.763	0.237	0,193	0,807	1
22	KITLG	rs642742	12	88905969	T	C	INTERGENIC	0.200	0.800	0.899	0.101	0,310	0,690	11
23	KITLG	rs12821256	12	88934558	T	C	REGULATORY	0.879	0.121	0.998	0.002	0,952	0,048	13,17
24	LYST	rs3768056	1	235744525	A	G	INTRONIC	0.772	0.228	0.914	0.086	0,795	0,205	19,20
25	MC1R	rs3212345	16	89915864	C	T	INTRONIC	0.424	0.576	0.942	0.058	0,543	0,457	12
26	MC1R	N29insA	16	89985753	DEL	InsA	EXONIC	-	-	-	-	0,997	0,002	13
27	MC1R	rs885479	16	89919746	G	A	EXONIC	0.930	0.070	0.993	0.007	0,902	0,098	13,17
28	MC1R	rs1805005	16	89919436	G	T	EXONIC	0.888	0.112	0.995	0.005	0,900	0,100	13,17
29	MC1R	rs1805006	16	89919510	C	A	EXONIC	0.990	0.010	0.999	0.001	0,992	0,008	13,17

30	MC1R	rs1805009	16	89920138	G	C	EXONIC	0.992	0.008	0.998	0.002	0,981	0,019	13,17
31	MC1R	rs1110400	16	89919722	T	C	EXONIC	0.992	0.008	1.000	0	0,997	0,003	13,17
32	MFSD12	rs10424065	19	3545024	C	T	INTRONIC	0.991	0.009	0.702	0.298	0,942	0,058	18
33	MFSD12	rs6510760	19	3565255	G	A	INTRONIC	0.938	0.062	0.164	0.836	0,499	0,501	18
34	MYO5A	rs1724630	15	52415803	G	C	INTRONIC	0.811	0.189	0.417	0.583	0,703	0,297	1
35	NPLOC4	rs9894429	17	81629785	C	T	SYNONYMOUS	0.588	0.412	0.185	0.815	0,453	0,547	19,20
36	OCA2	rs1800407	15	25903913	C	T	EXONIC	0.924	0.076	0.998	0.002	0,946	0,054	1,13,16,17
37	OCA2	rs7170989	15	27955262	C	T	INTRONIC	0.750	0.250	0.323	0.677	0,594	0,406	14
38	OCA2	rs1448484	15	28038295	G	A	INTRONIC	0.008	0.992	0.803	0.197	0,163	0,837	1
39	OCA2	rs1037208	15	27986211	T	G	INTRONIC	0.817	0.183	0.410	0.590	0,701	0,299	1
40	OCA2	rs1375164	15	28046666	T	C	INTRONIC	0.237	0.763	0.961	0.039	0,411	0,589	1
41	OCA2	rs1597196	15	28049776	T	G	INTRONIC	0.199	0.801	0.762	0.238	0,339	0,661	1
42	OCA2	rs1800404	15	27990627	C	T	SYNONYMOUS	0.214	0.786	0.873	0.127	0,390	0,610	1, 18
43	OCA2	rs1900758	15	27984951	C	T	INTRONIC	0.335	0.665	0.893	0.107	0,521	0,479	1
44	OCA2	rs2036213	15	27839034	C	A	INTRONIC	0.583	0.417	0.762	0.238	0,612	0,388	1
45	OCA2	rs2594935	15	27939892	A	G	INTRONIC	0.331	0.669	0.866	0.134	0,509	0,491	1
46	OCA2	rs3794606	15	28023862	C	A	INTRONIC	0.240	0.760	0.958	0.042	0,408	0,592	1
47	OCA2	rs4778137	15	28082689	G	C	INTRONIC	0.311	0.689	0.663	0.337	0,428	0,572	1
48	OCA2	rs4778138	15	28090674	G	A	INTRONIC	0.169	0.831	0.718	0.282	0,319	0,681	1
49	OCA2	rs4778241	15	28093567	A	C	INTRONIC	0.242	0.758	0.576	0.424	0,354	0,646	1
50	OCA2	rs4778232	15	28036619	T	C	INTRONIC	0.240	0.760	0.837	0.163	0,388	0,612	1
51	OCA2	rs895829	15	28052911	C	T	INTRONIC	0.245	0.755	0.962	0.038	0,400	0,600	1
52	OCA2	rs895828	15	28052887	C	G	INTRONIC	0.098	0.902	0.571	0.429	0,202	0,798	1
53	SCL24A4	rs12896399	14	92307319	G	T	INTERGENIC	0.570	0.430	0.979	0.021	0,661	0,339	1,13,16,17
54	SCL24A4	rs2402130	14	92334859	G	A	INTRONIC	0.213	0.787	0.600	0.400	0,232	0,768	13,17
55	SCL24A5	rs1426654	15	48134287	G	A	EXONIC	0.003	0.997	0.926	0.074	0,264	0,736	1,17,18
56	SCL45A2	rs16891982	5	33951588	C	G	EXONIC	0.062	0.938	0.964	0.036	0,350	0,650	1,13,16,17
57	SCL45A2	rs28777	5	33958854	C	A	INTRONIC	0.044	0.956	0.805	0.195	0,301	0,699	1,13,17
58	SCL45A2	rs13289	5	33986304	C	G	3' UTR	0.401	0.599	0.742	0.258	0,489	0,511	1
59	SCL45A2	rs183671	5	33964105	T	G	INTRONIC	0.047	0.953	0.790	0.210	0,302	0,698	21
60	TMEM138	rs7948623	11	61369675	A	T	REGULATORY	0.994	0.006	0.734	0.266	0,940	0,060	18
61	TTC3/DSC9	rs2835630	21	37149542	G	A	INTRONIC	0.476	0.524	0.654	0.346	0,534	0,466	19,20
62	TYR	rs1042602	11	89178528	C	A	EXONIC	0.628	0.372	0.988	0.012	0,670	0,330	1,13,17
63	TYR	rs1393350	11	89277878	G	A	INTRONIC	0.756	0.244	0.992	0.008	0,831	0,169	1,13,16,17

64	<i>TYRP1</i>	rs683	9	12709305	A	C	3' UTR	0.627	0.373	0.112	0.888	0,485	0,515	1,13,17
65	<i>TYRP1</i>	rs13289810	9	12396731	A	G	INTERGENIC	0.668	0.332	0.750	0.250	0,702	0,298	15
66	<i>TYRP1</i>	rs1325127	9	12668328	C	T	INTERGENIC	0.346	0.654	0.731	0.269	0,393	0,607	1
67	<i>TYRP1</i>	rs2733832	9	12704725	C	T	INTRONIC	0.428	0.572	0.946	0.054	0,558	0,442	1
68	<i>UGT1A6</i>	rs2070959	2	233693545	A	G	EXONIC	0.690	0.310	0.761	0.239	0,694	0,306	19,20

A1: ancestral allele; A2: derived allele; Except for Southern Brazil frequencies, data are available at <http://www.ensembl.org/index.html/>

MATERIALS AND METHODS

This project was approved by Pontifical Catholic University of Rio Grande do Sul Research Ethics Committee (CEP) under protocol #11-05722 (Of.CEP-0295/12 and Of.CEP-1041/12).

Samples and Phenotypical Characterization

An informed written consent and a survey form on sex, age, self-declared biogeographical ancestry, place of birth and residency was completed and signed by all voluntary participants.

Oral swabs were collected from 465 individuals from Southern Brazilian population.

Digital photographs of subjects' eyes were taken and their colors were classified as: Blue (112), Green (37), Hazel (33), Light Brown (45), or Dark Brown (238) (FIGURE 1). In addition, eyes' colors were grouped into five models (TABLE 2). Amounts of red (R), green (G), and blue (B) color values were measured by analyzing each photo with COLORS software [24] to confirm such classification. Eye color categorization was performed according to overall accepted perception for local Southern Brazilian population, and category assignment was independently performed by three different collaborators.

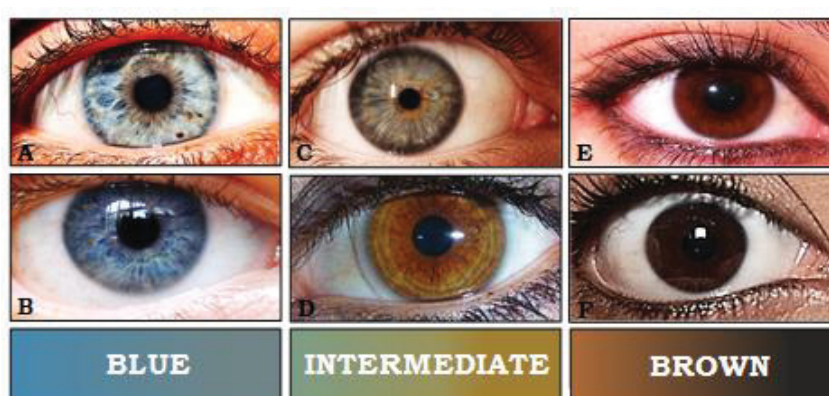


FIGURE 1 – Eye color phenotype in humans: Blue (A. Light Blue; B. Dark Blue), Intermediate (C. Green; D. Hazel), Brown (E. Light Brown; F. Dark Brown)

TABLE 2 – Eyes phenotypic classes grouping models

Model	Number of Classes	Phenotype Classes
3.1	3	Blue, Intermediate (Green, Hazel), Brown (Light and Dark)
3.2	3	Blue, Intermediate (Green, Hazel, Light Brown), Dark Brown
3.3	3	Light (Blue and Green), Intermediate (Hazel and Light Brown), Dark Brown
4	4	Blue, Green, Hazel, Brown (Light and Dark)
5	5	Blue, Green, Hazel, Light Brown, Dark Brown

Each participant skin color was identified using the Fitzpatrick Score (Types 1 to 6) and classified independently by at least three collaborators as: White (120), Pale (150), Beige (66), Light Brown (43), Medium Brown (59), or Dark Brown (27) (FIGURE 2). In addition, skin colors were grouped into six models (TABLE 3). Amounts of RGB color values were measured in an inner and hairless portion of the right arm (below elbow) using a color analyzer equipment ACR-1023 (Instrutherm, Brazil) [25-26].

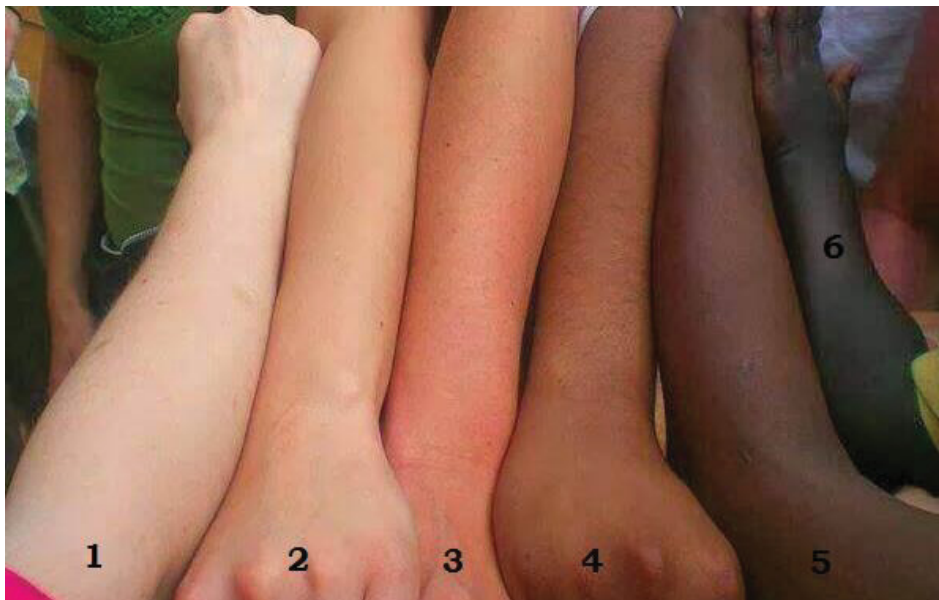
**FIGURE 2** – Skin color phenotype in humans: White (1), Pale (2), Beige (3), Light Brown (4), Medium Brown (5) and Dark Brown (6). [27]

TABLE 3 – Skin phenotypic classes grouping models

Model	Number of Classes	Phenotype Classes
2	2	Light (White, Pale, Beige), Dark (Light Brown, Medium Brown, Dark Brown)
3	3	Light (White, Pale), Intermediate (Beige, Light Brown), Dark (Medium Brown, Dark Brown)
4	4	Light (White, Pale), Beige, Light Brown, Dark (Medium Brown, Dark Brown)
5.1	5	Light (White, Pale), Beige, Light Brown, Medium Brown, Dark Brown
5.2	5	White, Pale, Beige, Light Brown, Dark (Medium Brown, Dark Brown)
6	6	White, Pale, Beige, Light Brown, Medium Brown, Dark Brown

DNA Genotyping

Genomic DNA from oral swabs was extracted using the NucleoSpin Tissue kit (Macherey-Nagel Inc.) following manufacturer instructions or by organic extraction method [28]. Samples were sent to Liggins Institute, from Auckland University, New Zealand, for SNPs genotyping using iPLEX® Pro reagentes on the MassARRAY® system and MassARRAY Typer application software [29-30].

Statistical Analysis

Bioinformatic instruments were used to analyze collected data. Two distinct Feature Selection (FS) techniques were used to identify the most relevant SNPs to describe a given class: Recursive Feature Elimination (RFE) [31-32] and Chi-Squared Test (χ^2) [33]. Three alternative Machine Learning (ML) algorithms were experimented to construct predictive models based in identified SNPs in FS step: Support Vector Machine (SVM) [34], Random Forest (RF) [35], and Multi-Layer Perceptron (MLP) [36-37].

A ML algorithm infers/generalizes properties of a given input set, and that information allows it to make predictions about other data that it might see in the future. In Big Data era, both sample size and feature numbers are continuously growing over time [38]. This large amount of information can reduce ML techniques ability to

generalize data. Training an algorithm with irrelevant features will affect performance of proposed predictive models. FS is a task to select the minimum number of attributes needed to represent data accurately [38]. Relevant features use can improve ML predictive accuracy, reducing learning period, and providing a simple model that better explains data [38]. Feature Selection is an essential part of building Machine Learning models. FS methods are applied in the most different domains such as gene selection in cancer research, text mining, bank data analysis, and forensics. It is possible to derive a general architecture from most feature selection algorithms. It consists of four basic steps: subset generation, subset evaluation, stopping criterion, and result validation [39]. Feature selection algorithms create a subset, evaluate it, and look until an ending criterion is satisfied. Finally, subset found is validated by classifier algorithm on real data. Using as few features as possible also reduces ML models complexity, which means less time and computer power are needed to simulation runs and model overall comprehension, with understanding of individual elements contribution, is easier to achieve.

RESULTS

Eye Color

From RFE and X^2 feature selectors, a search was made for SNPs best suitable to discriminate eye phenotypic classes. Then, three ML algorithms (SVM, RF and MLP) were applied to construct predictive models based in SNPs identified in FS step (FIGURE 3). Several SNPs selections were analyzed for all five class-grouping models. The best phenotypic classification was found using ten SNPs, using RFE feature selection and RF ML algorithm, in 3.1 classes model, which had an overall accuracy of 86% in 10-fold cross validation (FIGURE 4). Selection of another number of SNPs by RFE or X^2 , with respective machine learning predictive accuracy by SVM, RF and MLP can be verified in supplementary material (FIGURE S1-S5 and S12-S16; TABLE S1-S2 and S7-S11). Since this model was found to achieve better accuracy results, it was chosen to be employed in the eye color phenotype predictor system. Besides, the use of this particular categories model, including three phenotypic categories distributed as described, seems to reflect better visual grouping of different eye colors, since different shades of brown might be visually undistinguishable at naked eye or at distant inspection, as well as noticeable similarities between green and hazel eyes, where positive categorization might be difficult. This fact has particular relevance considering the forensic application aspect to which proposed SNP panels are designed to contribute.

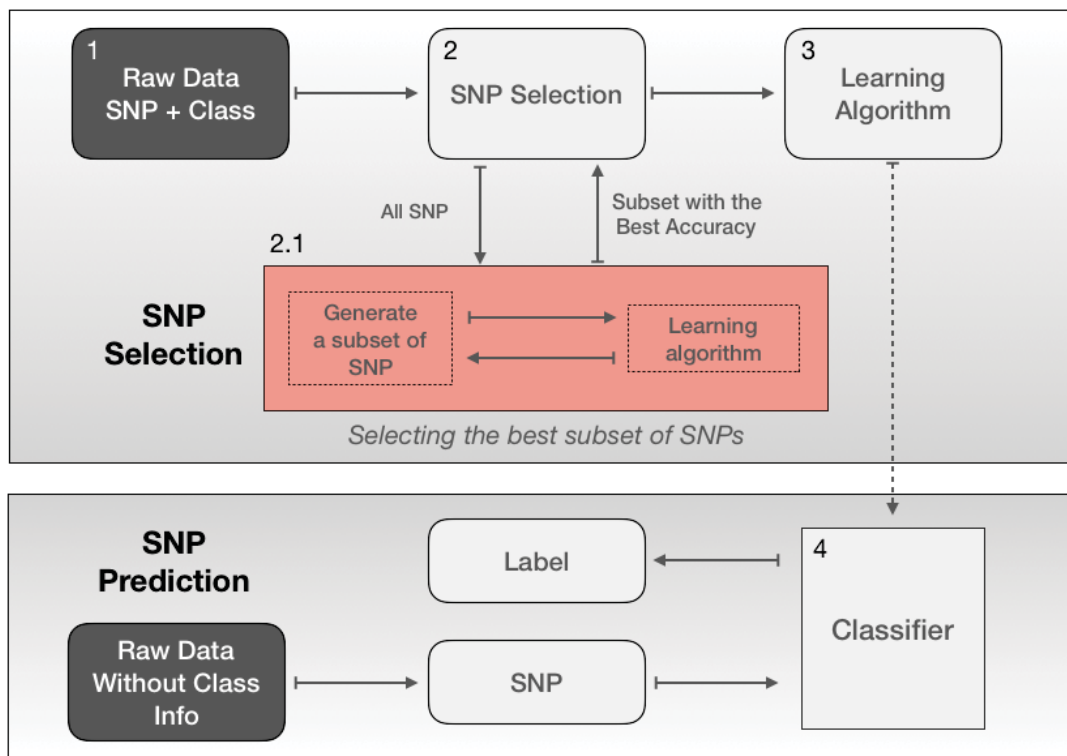


FIGURE 3 – Analysis methodology pipeline. 1. Input data (SNPs and phenotype classes); 2. SNP selection by RFE or X2; 3. Machine learning process by SVM, RF or MLP; 4. Classifier construction.

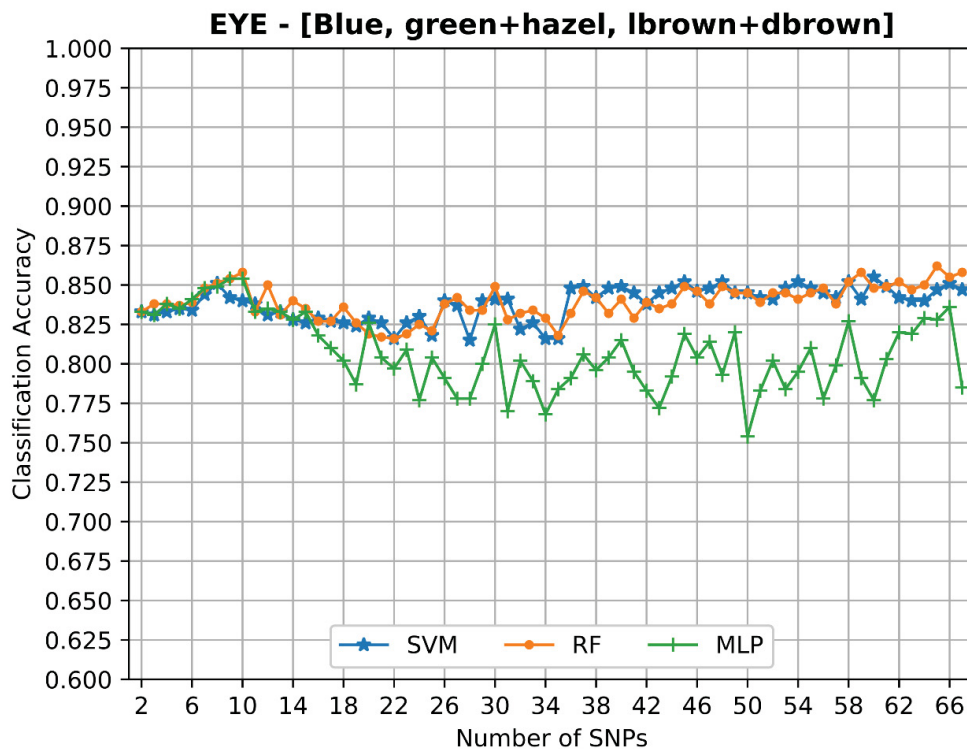


FIGURE 4 – Accuracy classification in accordance with SNPs selection by RFE and SVM / RF / MLP – EYE – 3.1 CLASSES: Blue; Intermediate (Green / Hazel); and Brown (Light / Dark Brown).

In this model, 431 samples were considered, and the ten selected SNPs included in proposed model, and their associated genes, were: *ASIP* (rs6119471), *DDB1* (rs11230664), *HERC2* (rs6497271, rs1129038, rs12913832), *MFSD12* (rs10424065, rs6510760), *OCA2* (rs1800407), *SLC24A5* (rs1426654) and *SLC24A4* (rs2402130). FIGURE 5 shows data clustering, based on ten SNPs, and phenotypic classes overlap.

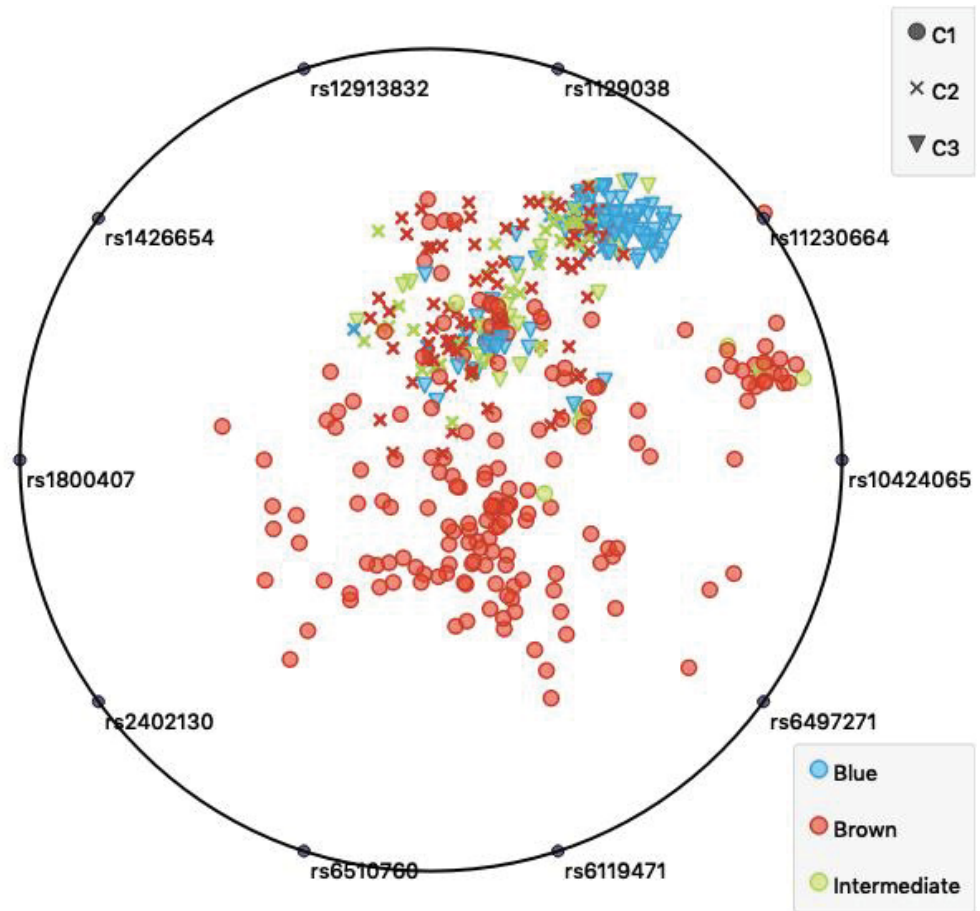


FIGURE 5 – Data clustering (C1, C2 and C3), based on ten SNPs, and eyes phenotypic classes (Blue, Intermediate and Brown) overlap.

SNPs act together for phenotype definition. However, some have greater segregation power than others. Some SNPs individual contribution to phenotype can be verified in FIGURE 6.

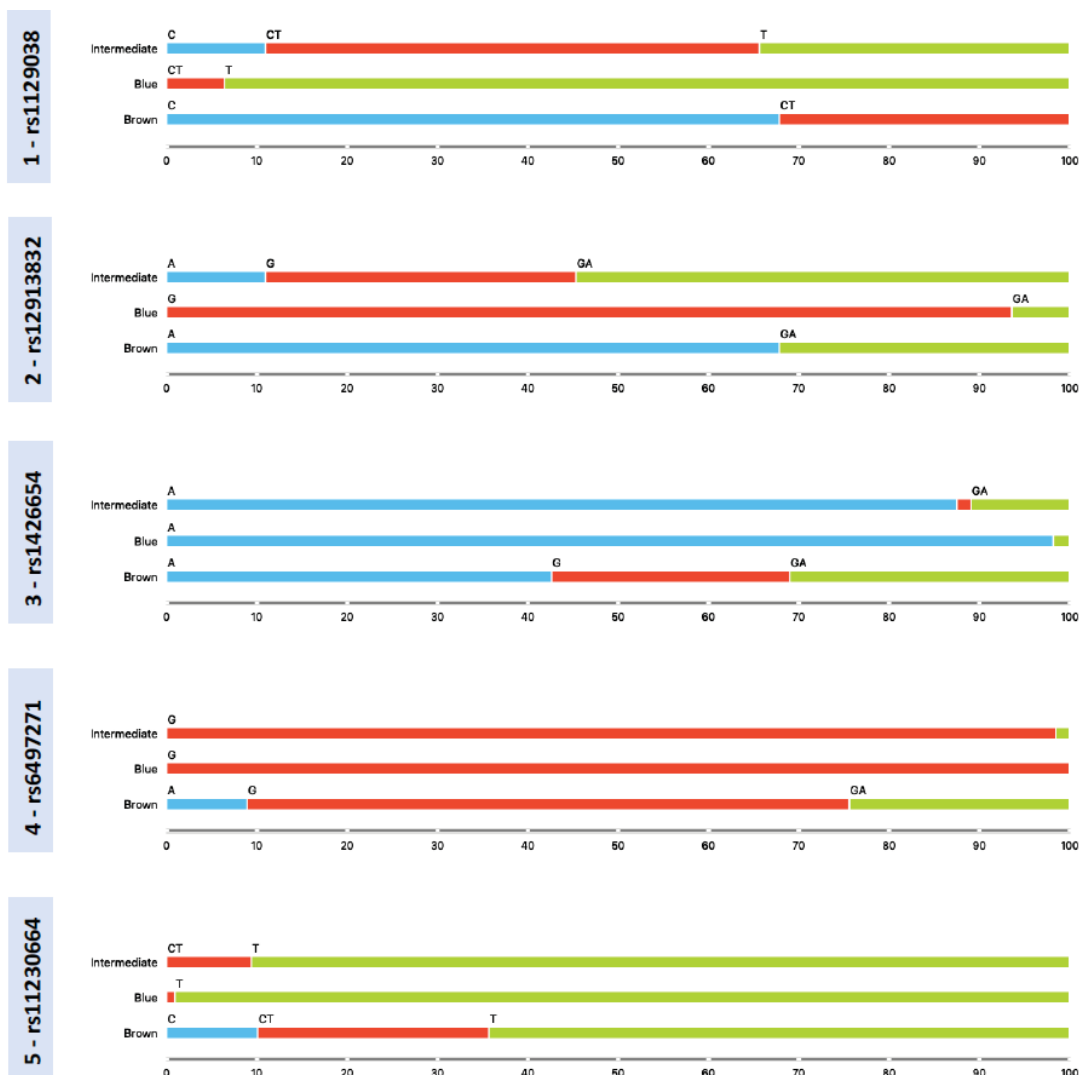


FIGURE 6 – Five best SNPs for eyes color prediction, with individual contribution, considering three phenotypic classes (Blue, Intermediate and Brown, model 3.1), in descending importance order. Classification power of remaining five SNPs can be seen in the supplementary material (FIGURE S23A-B).

One of the 10-folds cross validation was chosen as basis to the classifier construction, increasing predictor overall accuracy to 87%. Sensitivity in predicting brown eyes was 98%, and blue eyes was 94%, whereas intermediate eyes color sensitivity was significantly lower. Eyes classifier confusion matrix is shown in FIGURE 7, and prediction statistics in TABLE 4.

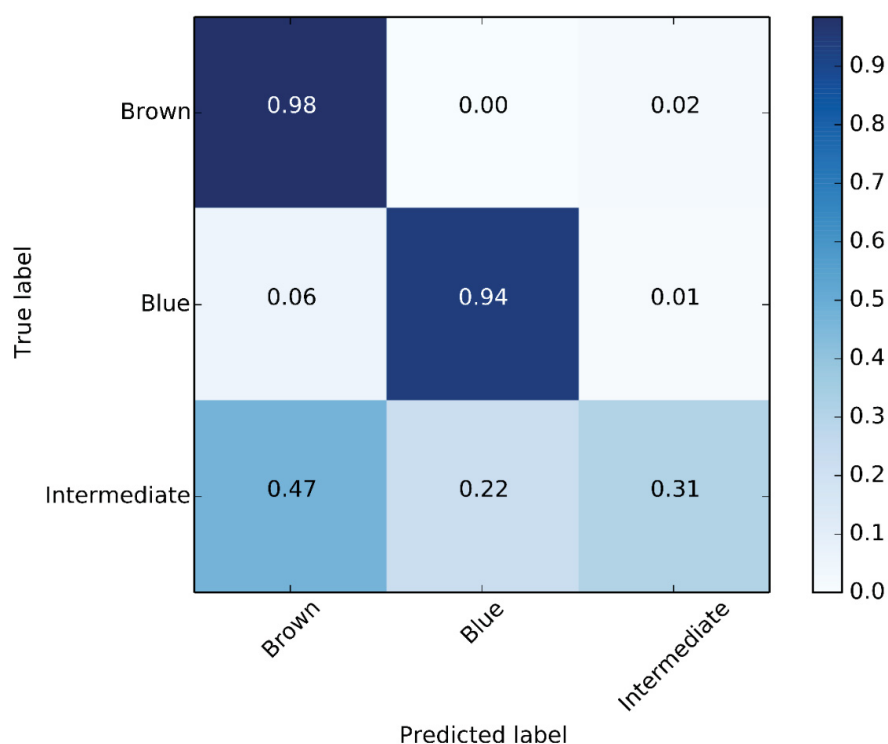


FIGURE 7 – Eyes color classification in three classes (Blue, Intermediate and Brown) by the predictor.

TABLE 4 – Eyes and skin phenotype prediction statistics, for best observed models.

Statistics	Eyes			Skin		
	Blue	Intermediate	Brown	Light	Intermediate	Dark
Sensitivity	93.57	31.25	98.45	100.00	90.48	97.67
Specificity	95.65	98.64	79.19	94.49	99.65	99.69
PPV	87.93	80.00	87.58	97.20	98.70	97.67
NPV	97.78	89.16	97.16	100.00	97.27	99.69
Accuracy	95.13	88.63	90.72	98.11	97.57	99.46

Values in percentage (%), Positive predictive value (PPV) = Precision, Negative predictive value (NPV)

Skin Color

From RFE and X^2 feature selectors, a search was made for SNPs best suitable to discriminate skin phenotypic classes. Then, three ML algorithms (SVM, RF and MLP) were applied to construct predictive models based in SNPs identified in FS step (FIGURE 3). Several SNPs selections were analyzed for all six class-grouping models. We choose the phenotypic selector with twenty five SNPs, using RFE feature selection and SVM learning algorithm, in 3 classes model, which had an overall accuracy of 81% in 10-fold cross validation (FIGURE 8). Selection of another number of SNPs by RFE or X^2 , with respective machine learning accuracy by SVM, RF and MLP can be verified in supplementary material (FIGURE S6-S11 and S17-S22; TABLE S3-S6 and S12-S17).

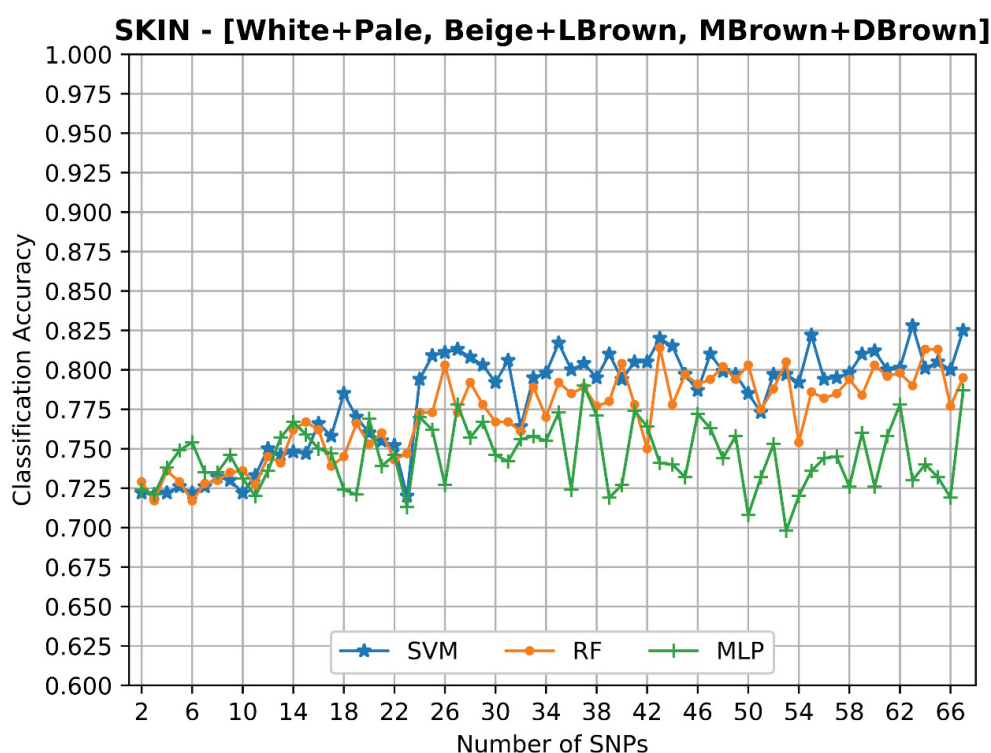


FIGURE 8 – Accuracy classification in accordance with SNPs selection by RFE and SVM / RF / MLP – SKIN – 3 CLASSES: Light (White / Pale); Intermediate (Beige / Light Brown); Dark (Medium Brown / Dark Brown).

In this model, 370 samples were considered, and 25 selected SNPs included in the best overall performance model were as follows, including their associated gene: *ASIP* (rs2424984, rs6119471), *DDB1* (rs11230664), *HERC2* (rs1129038, rs11636232, rs16950987, rs2238289, rs4932620, rs6497271), *MC1R* (rs1805006, rs1805009, rs3212345, rs885479), *MFSD12* (rs10424065, rs6510760), *NPLOC4* (rs9894429), *OCA2* (rs1037208, rs1375164, rs3794606, rs4778232), *SLC24A5* (rs1426654), *SLC45A2* (rs16891982, rs183671), *TMEM138* (rs7948623) and *TYR* (rs1042602). FIGURE 9 shows data clustering, based on 25 SNPs, and phenotypic classes overlap.

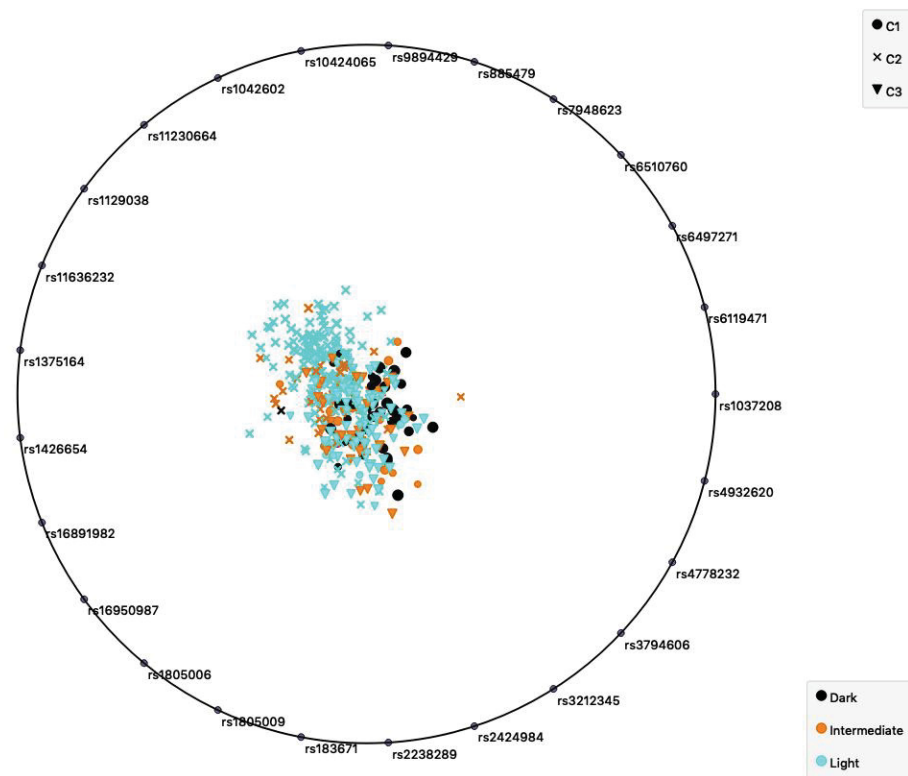


FIGURE 9 – Data clustering (C1, C2 and C3), based on 25 SNPs, and skin phenotypic classes (Light, Intermediate and Dark) overlap.

SNPs act together for phenotype definition. In skin case, more SNPs are needed to achieve appropriate classes identification. Some SNPs individual contribution to phenotype can be verified in FIGURE 10.

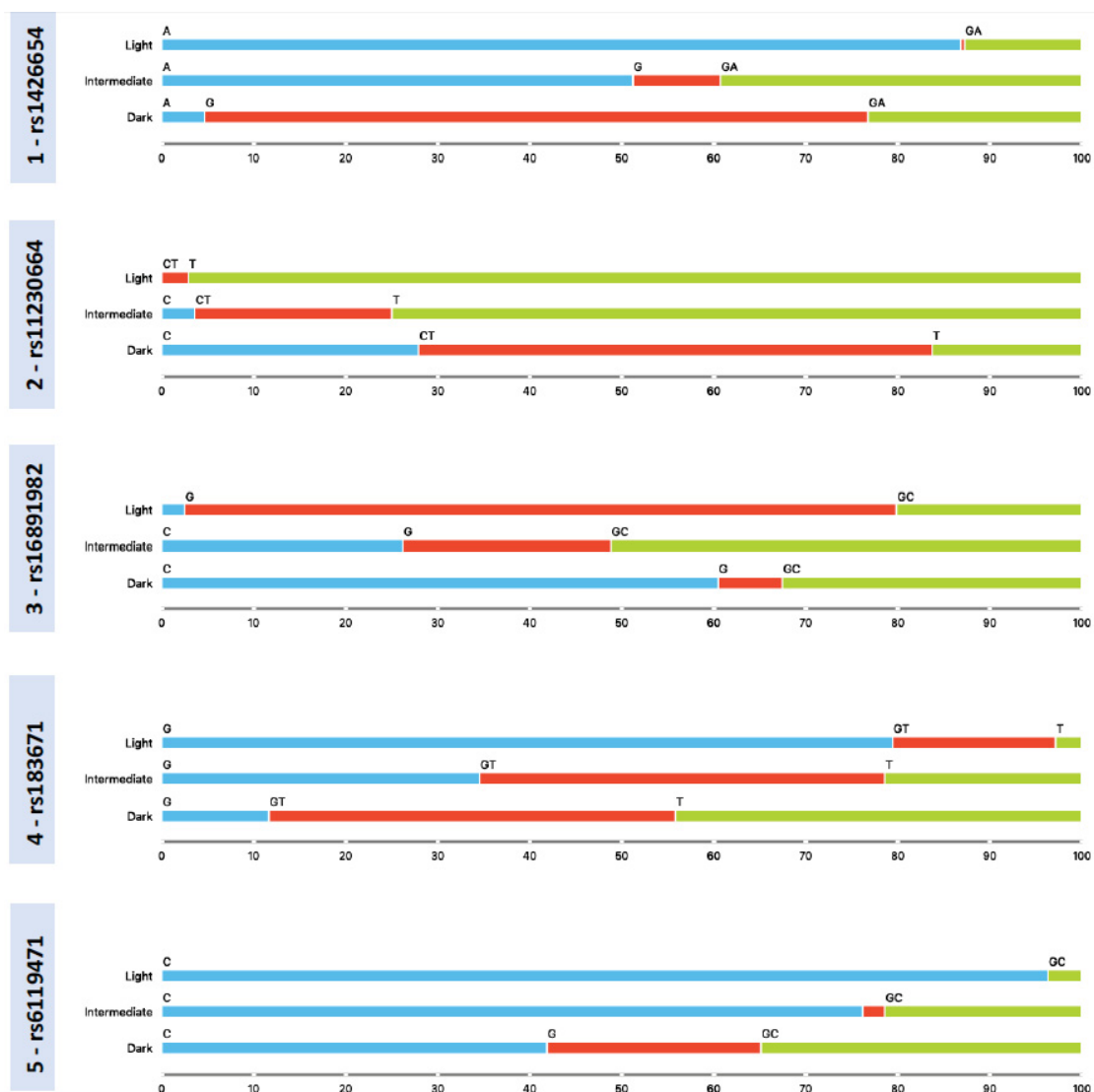


FIGURE 10 – Five best SNPs for skin color prediction, with individual contribution, considering three phenotypic classes (Light, Intermediate and Dark), in descending relevance order. Classification power of remaining 20 SNPs can be visualized in the supplementary material (FIGURE S24A-E).

Once again, one of the 10-folds cross validation was selected to construct the classifier, increasing predictor overall accuracy to 97,5%. This may be due to the fact that the classes are unbalanced and, depending on the grouping performed in the simulation, overall obtained accuracy can increase or decrease. Sensitivity in predicting light skin was 100%, in dark skin was 98%, and in intermediate skin was 90%. Skin classifier confusion matrix is shown in FIGURE 11, and prediction statistics in TABLE 4.

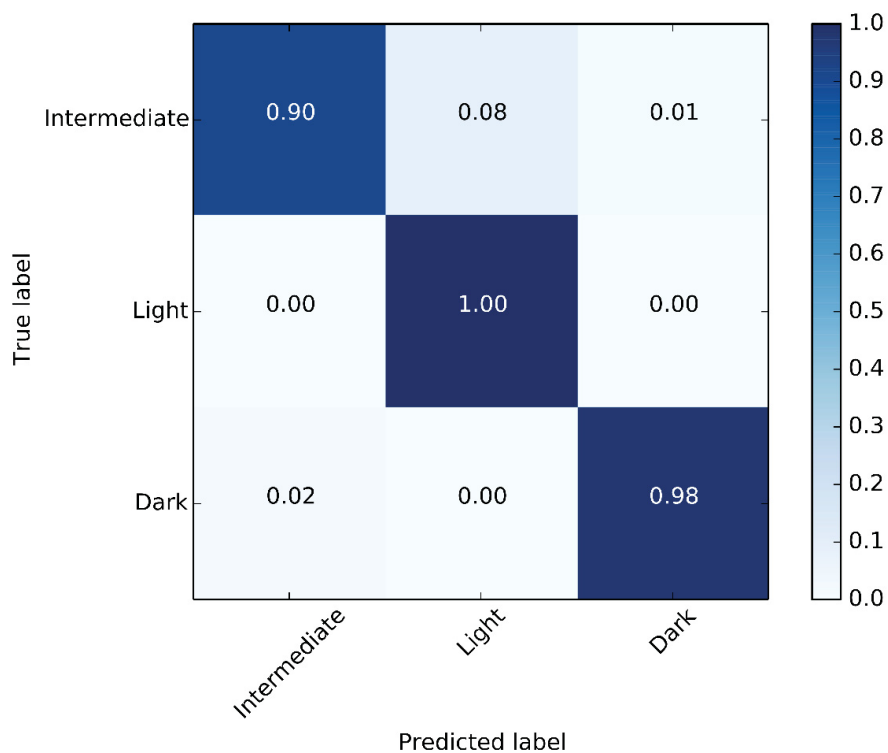


FIGURE 11 – Skin color classification in three classes (Light, Intermediate and Dark) by the predictor.

DISCUSSION

As can be seen in FIGURE 4 graphic, ten selected SNPs with RF technique were sufficient to segregate three eyes phenotypic classes (blue, intermediate and brown). Additional SNPs inclusion in prediction models leads to noticeable accuracy decrease. RF and SVM techniques show more consistent results in considered models, whereas MLP technique shows highly variable accuracy depending on selected SNPs number. MLP shows increased sensitivity than alternative evaluated algorithms when analyzing data comprised of unbalanced classes [36]. This particular three phenotype classes model is also the most generally used worldwide in prediction models, segregating blue and brown eyes, extremes in regard to melanin pigmentation, from eyes with intermediate colors (green and hazel) [13,40], often more difficult to define phenotypically and genotypically.

Human iris varies from lighter shades of blue to dark brown and this difference is determined by the type, amount and distribution of melanin in iris outer layer, which, in turn, are directly or indirectly influenced by several genes and their alleles [41]. Blue eyes have low concentration of both melanin and melanosomes, while green eyes contain moderate melanin and melanosomes levels, and brown eyes have higher pigment levels and large amount of melanosomes [41]. Different combinations in concentration of two types of melanin are responsible for different observable shades: pheomelanin, a reddish-yellow pigment, generally features lighter colorations and eumelanin, usually present in darker colorations [42]. Blue and green eyes occurrence is practically limited to ethnical Europeans, while brown eyes are found in most of the human population throughout the world [43].

Genotype-phenotype associations found here were also observed in previous studies conducted in different populations. *HERC2* gene, located near *OCA2*, is the most important eye color-related gene and regulates *OCA2* expression [9-10,43]. SNPs rs1129038 and rs12913832 were the best for segregating blue eyes from other colors. rs12913832 is considered the best SNP for predicting eye color [1,9-10,13,16-17,43-45,46]. Ruiz *et al.* suggested that adding additional *HERC2* SNPs would increase prediction accuracy: rs1129038 improves eye color prediction in general [47]. In this model, rs6497271, a SNP selected from an article that studied skin color in Africans, also promoted accuracy increase in eyes color prediction [18]. Third best SNP to identify classes in this model is rs1426642, in agreement with previous studies

[1,17,18,40,46]. Other SNPs have smaller contributions and act together to determine pigmentation: rs11230664 [18], rs6119471 [1,17], rs10424065 [18], rs2402130 [13,17]. rs1800407 was associated with intermediate color (green and hazel) [1,13,16-17,48-49].

As can be seen in FIGURE 8 graphic, the highest accuracies were found by selecting 25 (0.81), 35 (0.82) or 63 SNPs (0.83) by SVM technique. We choose the 25 SNPs model, because we thought it would not be worth adding 10 snps the genotyping procedure in order to obtain an increase of only 1% in overall model accuracy, or adding 38 SNPs to increase accuracy by 2%. We were successful, despite presence of unbalanced classes in sample, in finding simulations with high general accuracy and high sensitivity for all colors (light, intermediate and dark). This three phenotype classes model, derived from Fitzpatrick Score, is an interesting simplification because it groups light skin phenotypes (types 1 and 2) and dark skin phenotypes (types 5 and 6), although intermediate skin group is admissible as more heterogeneous (types 3 and 4). Just like eyes intermediate color, precise skin intermediate color classification is significantly more challenging than its extreme counterparts [17], but we achieved 90% of correct classification in this model.

There is a correlation between climatic distribution, ultraviolet radiation (UV) levels and human skin pigmentation, which influenced genes mutations and new alleles appearance. Equator line has higher UV incidence and people living in this region generally have darker skin to protect themselves from it, while populations at high latitudes, with less UV radiation, have less pigmented skin to provide vitamin D production [50]. Individuals with light or dark skin have similar melanocytes numbers but those with light skin possess less pigmented melanosomes, rich in pheomelanin, and dark skin individuals have larger, more numerous and pigmented melanosomes [51].

Selection of these 25 variables was interesting given that genotype-phenotype associations found here were also observed in previous studies conducted in different populations. Liu *et al.* suggested that rs1426654 (first most important SNP in this model) and rs16891982 (third most important SNP) are likely major predictors for skin color differences observed between continents [40] as well as Sawitzki *et al.* [46], Maroñas *et al.* [52], Spichenok *et al.* [53], and Mushailov *et al.* [54]. Crawford *et al.*, in a study of African population, found association of rs11230664 (second most important SNP in this model) with skin pigmentation, as well as rs6497271 (sixth better SNP),

rs7948623 (tenth SNP), rs10424065 (fourteenth SNP), rs4932620 (twentieth SNP) and rs1426654 [18]. Chaitanya *et al.* and Walsh *et al.* found association between rs1426654, rs16891982, rs6119471, rs1129038, rs1042602, rs2238289, rs8854479, rs1805006 and rs1805509 with skin pigmentation [17,55]. Adhikari *et al.*, found association of rs183671 (fourth most importante SNP in this model) with skin color [21].

SNPs act together for phenotype definition. In skin example, more SNPs are needed to classes identification, and most of them have a small participation in skin color definition. Other SNPs found were: rs1375164, rs3794606, rs4778232, rs2424984, rs11636232, rs1037208 [1]; rs16950987 [8]; rs3212345 [12]; and rs9894429 [19-20].

rs6510760, although selected by eyes and skin predictors, did not segregate any group, as all individuals were heterozygotes GA. As we select the number of SNPs that contribute to the prediction, probably this one that did not interfere with the accuracy was added to complete the selection. Therefore, its absence in a proposed phenotype prediction panel can be regarded as indifferent to resulting prediction power.

One important thing about our DNA phenotyping system is that it is very versatile, since accuracy in most models is shown to display low variation levels, so any of them (points in graphics) can be selected to create a predictor. Antique or degraded samples, which may not be able to originate a complete profile due to lack of fragment amplification during genotyping, can use a different predictor with fewer SNPs (two SNPs for eyes prediction already guarantees a general accuracy of at least 83%, for example).

BR-EASE beta version, is an easy-to-use system, suitable for adoption by researchers, law enforcement agencies and general forensic community as an aid in investigations and as a teaching resource of multiple gene inheritance. This system predictor is available at '<http://sbcdb.inf.ufrgs.br/brease>'.

CONCLUSION

Forensic DNA phenotyping have been showing major technological advances and is just steps away to be admitted into forensic labs casework routine, despite ethical concerns, with great applicability and reliability. More studies are needed to better distinguish intermediate phenotype classes, not only from their extreme counterparts, but even in variant color shades classified inside this same category. In this work, we introduced a simple system for eyes and skin phenotype prediction through DNA: BR-EASE, Brazilian Eye and Skin Estimator. We believe this FDP system can contribute to significant advances in forensic science, not only in a local level, but through nationwide application of its principles by law enforcement agencies, leading to a more efficient and reliable criminal justice system in Brazil.

FINANTIAL SUPPORT

This work was supported by CAPES Brazil (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Edital 25/2014 – Pró-Forenses), CNPq Brazil (Conselho Nacional de Pesquisa Científica) and FAPERGS (Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul - Chamada nº 16/2014 – INCT).

ACKNOWLEDGMENTS

The authors would like to thank Deborah Soares Bispo Santos Silva and Pietra Graebin for technical support.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] M. Kayser, Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes, *Forensic Sci. Int. Genet.* 18 (2015) 33-48. <http://doi.org/10.1016/j.fsigen.2015.02.003>.
- [2] F. Liu, K. van Duijn, J.R. Vingerling, A. Hofman, A.G. Uitterlinden, A.C.J.W. Janssens, M. Kayser, Eye color and the prediction of complex phenotypes from genotypes, *Curr. Biol.* 19 (2009) R192–R193. <http://doi.org/10.1016/j.cub.2009.01.027>.
- [3] M. Kayser, P.M. Schneider, DNA-based prediction of human externally visible characteristics in forensics, motivations, scientific challenges, and ethical considerations, *Forensic Sci. Int. Genet.* 3 (2009) 154-161. <http://doi.org/10.1016/j.fsigen.2009.01.012>.
- [4] A.J. Brookes, The essence of SNPs, *Gene.* 234 (1999) 177-186. [https://doi.org/10.1016/S0378-1119\(99\)00219-X](https://doi.org/10.1016/S0378-1119(99)00219-X).
- [5] B. Budowle, SNP Typing Strategies, *Forensic Sci. Int.* 146 (2004) S139-S142. <https://doi.org/10.1016/j.forsciint.2004.09.042>.
- [6] R. Javed and Mukesh, Current research status, databases and application of single nucleotide polymorphism, *Pakistan Journal of Biological Sciences.* 13 (2010) 657-663. <https://doi.org/10.3923/pjbs.2010.657.663>.
- [7] L.C. Jacobs, A. Wollstein, O. Lao, A. Hofman, C.C. Klaver, A.G. Uitterlinden, T. Nijsten, M. Kayser, F. Liu, Comprehensive candidate gene study highlights UGT1A and BNC2 as new genes determining continuous skin color variation in Europeans. *Hum Genet.* 132 (2013) 147-158. <https://doi.org/10.1007/s00439-012-1232-9>.
- [8] P. Sulem, D.F. Gudbjartsson, S.N. Stacey, A. Helgason, T. Rafnar, M. Jakobsdottir, S. Steinberg, S.A. Gudjonsson, A. Palsson, G. Thorleifsson, S. Pálsson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K.R. Benediktsdottir, K.K. Aben, S.H.

Vermeulen, A.M. Goldstein, M.A. Tucker, L.A. Kiemeny, J.H. Olafsson, J. Gulcher, A. Kong, U. Thorsteinsdottir, K. Stefansson, Two newly identified genetic determinants of pigmentation in Europeans. *Nat Genet* 40 (2008) 835-837. <https://doi.org/10.1038/ng.160>.

[9] R.A. Sturm, D.L. Duffy, Z.Z. Zhao, F.P.N. Leite, M.S. Stark, N.K. Hayward, N.G. Martin, G.W. Montgomery, A single SNP in an evolutionary conserved region within intron 86 of the *HERC2* gene determines human blue-brown eye color. *Am. J. Hum. Genet.* 82 (2008) 424-431. <https://doi.org/10.1016/j.ajhg.2007.11.005>.

[10] H. Eiberg, J. Troelsen, M. Nielsen, A. Mikkelsen, J. Mengel-From, K.W. Kjaer, L. Hansen, Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the *HERC2* gene inhibiting *OCA2* expression. *Hum. Genet.* 123 (2008) 177-187. <https://doi.org/10.1007/s00439-007-0460-x>.

[11] R.A. Sturm, Molecular genetics of human pigmentation diversity. *Hum. Mol. Genet. Review.* 18 (2009) R9-R17. <https://doi.org/10.1093/hmg/ddp003>.

[12] V. Neitzke-Montinelli, T.P. Urmenyi, E. Rondinelli, P.H. Cabello, R. Silva, R.S. Moura-Neto, Polymorphisms upstream of the melanocortin-1 receptor coding region are associated with human pigmentation variation in a Brazilian population. *Am. J. Hum. Biol.* 24(6) (2012):853-855. <https://doi.org/10.1002/ajhb.22301>.

[13] S. Walsh, F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, W. Branicki, M. Kayser, The HirisPlex system for simultaneous prediction of hair and eye colour from DNA, *Forensic. Sci. Int. Genet.* 7 (2013) 98-115. <https://doi.org/10.1016/j.fsigen.2012.07.005>.

[14] V. Kastelic, K. Drobic, Single multiplex system of twelve SNPs: Validation and implementation for association of SNPs with human eye and hair color. *Forensic Sci. Int. Genet.* 3 (2011) e216-e217. <https://doi.org/10.1016/j.fsigs.2011.08.108>.

- [15] E.E. Kenny, N.J. Timpson, M. Sikora, M. Yee, A.M. Estrada, C. Eng, S. Huntsman, E.G. Burchard, M. Stoneking, C.D. Bustamante, S.M. Myles, Melanesians blond hair is caused by an amino acid change in TYRP1. *Science* (2012) 336(6081):554. <http://doi.org/10.1126/science.1217849>.
- [16] S. Walsh, A. Lidenbergh, S.B. Zuniga, T. Sijen, P. Knijff, M. Kayser, K.N. Ballantyne, Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence, *Forensic Sci. Int. Genet.* 5 (2011) 464-471. <http://doi.org/10.1016/j.fsigen.2010.09.008>.
- [17] L. Chaitanya, K. Breslin, S. Zuñiga, L. Wirken, E. Pośpiech, M. Kukla-Bartoszek, T. Sijen, P. de Knijff, F. Liu, W. Branicki, M. Kayser, S. Walsh, The HirisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation, *Forensic Sci. Int. Genet.* 35 (2018) 123–135. <https://doi.org/10.1016/j.fsigen.2018.04.004>.
- [18] N.G. Crawford, D.E. Kelly, M.E.B. Hansen, M.H. Beltrame, S. Fan, S.L. Bowman, E. Jewett, A. Ranciaro, S. Thompson, Y. Lo, S.P. Pfeifer, J.D. Jensen, M.C. Campbell, W. Beggs, F. Hormozdiari, S.W. Mpoloka, G.G. Mokone, T. Nyambo, D.W. Meskel, G. Belay, J. Haut, NISC Comparative Sequencing Program†, H. Rothschild, L. Zon, Y. Zhou, M.A. Kovacs, M. Xu, T. Zhang, K. Bishop, J. Sinclair, C. Rivas, E. Elliot, J. Choi, S.A. Li, B. Hicks, S. Burgess, C. Abnet, D.E. Watkins-Chow, E. Oceana, Y.S. Song, E. Eskin, K.M. Brown, M.S. Marks, S.K. Loftus, W.J. Pavan, M. Yeager, S. Chanock, S. Tishkoff, Loci associated with skin pigmentation identified in African populations. *Science*. (2017) 358(6365): 8433 <https://doi.org/doi:10.1126/science.aan8433>.
- [19] F. Liu, A. Wollstein, P.G. Hysi, G.A. Ankra-Badu, T.D. Spector, D. Park, G. Zhu, M. Larsson, D.L. Duffy, G.W. Montgomery, D.A. Mackey, S. Walsh, O. Lao, A. Hofman, F. Rivadeneira, J.R. Vingerling, A.G. Uitterlinden, N.G. Martin, C.J. Hammond, M. Kayser, Digital Quantification of Human Eye Color Highlights Genetic Association of Three New Loci. *PLoS Genet.* 6 (2010) e1000934. <https://doi.org/10.1371/journal.pgen.1000934>.

- [20] A. Wollstein, S. Walsh, F. Liu, U. Chakravarthy, M. Rahu, J.H. Seland, G. Soubrane, L. Tomazzoli, F. Topouzis, J.R. Vingerling, J. Vioque, S. Böhringer, A.E. Fletcher, M. Kayser. Novel quantitative pigmentation phenotyping enhances genetic association, epistasis, and prediction of human eye colour. *Scientific Reports*. 7 (2017) 43359. <https://doi.org/10.1038/srep43359>.
- [21] K. Adhikari, T. Fontanil, S. Cal, J. Mendoza-Revilla, M. Fuentes-Guajardo, J.C. Chacón-Duque, F. Al-Saadi, J.A. Johansson, M. Quinto-Sanchez, V. Acuña-Alonzo, C. Jaramillo, W. Arias, R.B. Lozano, G.M. Pérez, J. Gómez-Valdés, H. Villamil-Ramírez, T. Hunemeier, V. Ramallo, C.C.S. de Cerqueira, M. Hurtado, V. Villegas, V. Granja, C. Gallo, G. Poletti, L. Schuler-Faccini, F.M. Salzano, M.C. Bortolini, S. Canizales-Quinteros, F. Rothhammer, G. Bedoya, R. Gonzalez-José, D. Headon, C. López-Otín, D.J. Tobin, D. Balding, A. Ruiz-Linares. A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. *Nature Communications* 7 (2016) 10815 <https://doi.org/10.1038/ncomms10815>.
- [22] T. N. Frudakis, *Molecular photofitting, Predicting Ancestry and Phenotype Using DNA* Elsevier, 2008.
- [23] T.C. Lins, R.G. Vieira, B.S. Abreu, D. Grattapaglia, R.W. Pereira, Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs, *Am. J. Hum. Biol.* 22 (2010) 187-92. <https://doi.org/10.1002/ajhb.20976>.
- [24] I. Otaka, K. Kumagai, Y. Inagaki, M. Shimoyama, K. Saegusa, T. Hara, Simple and inexpensive software designed for the evaluation of color, *Am. J. Ophthalmol.* 133 (2002) 140-142. [https://doi.org/10.1016/S0002-9394\(01\)01213-2](https://doi.org/10.1016/S0002-9394(01)01213-2).
- [25] F.C. Parra, R.C. Amado, J.R. Lambertucci, J. Rocha, C.M. Antunes, S.D. Pena, Color and genomic ancestry in Brazilians, *Proc. Natl. Acad. Sci. USA*. 100 (2003) 177-182. <https://doi.org/10.1073/pnas.0126614100>.

- [26] T.B. Fitzpatrick, The validity and practicality of sun-reactive skin types I through VI, *Arch. Dermatol.* 124 (1988) 869–871. <https://doi.org/10.1001/archderm.1988.01670060015008>.
- [27] Available at <https://geneticliteracyproject.org/2016/06/13/ancestral-implications-differing-skin-color/> (accessed in March 05, 2019).
- [28] D. Liu, *Handbook of Nucleic Acid Purification*. (2009).
- [29] Protocols available at '<https://agenacx.com/wp-content/uploads/2017/12/iPLEX-Pro-Reagents-User-Guide-1.pdf>'
- [30] S. Gabriel, L. Ziaugra, D. Tabbaa. SNP Genotyping Using the Sequenom MassARRAY iPLEX Platform. *Current Protocols in Human Genetics*. 60 (2009) 2.12.1-2.12.18. <https://doi.org/10.1002/0471142905.hg0212s60>.
- [31] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, (2002) 389-422. <https://doi.org/10.1023/A:1012487302797>.
- [32] I. Guyon and A. Elisseeff, An introduction to variable and feature selection. *Journal of Machine Learning Research* 3 (2003) 1157-1182.
- [33] N. Pandis. The chi-square test. *Am. J. Orthod. Dentofacial Orthop.* 150 (2016) 898-899. <https://doi.org/10.1016/j.ajodo.2016.08.009>.
- [34] C. Cortes and V. Vapnik, Support-vector networks. *Mach Learn* 20 (1995) 273-297. <https://doi.org/10.1007/BF00994018>.
- [35] T.K. Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 20 (1998) 832-844. <https://doi.org/10.1109/34.709601>.

- [36] F. Rosenblatt, *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan Books, Washington DC, 1961.
- [37] J.M. González-Camacho, J. Crossa, P. Pérez-Rodríguez, L. Ornella, D. Gianola, Genome-enabled prediction using probabilistic neural network classifiers. *BMC Genomics* 17 (2016) 208. <https://doi.org/10.1186/s12864-016-2553-1>.
- [38] J. Li, K. Cheng, S. Wang, F. Morstatter, R.P. Trevino, J. Tang, H. Liu, Feature selection: A data perspective. *ACM Comput. Surv.* 50 (2017) 94:1-94:45. <https://doi.org/10.1145/3136625>.
- [39] M. Dash, H. Liu, Feature selection for classification. *Intelligent Data Analysis*. 1 (1997) 131-156. [https://doi.org/10.1016/S1088-467X\(97\)00008-5](https://doi.org/10.1016/S1088-467X(97)00008-5).
- [40] F. Liu, B. Wenb, M. Kayser, Colorful DNA polymorphisms in humans, *Semin. Cell. Dev. Biol.* 24 (2013) 562-575. <http://doi.org/10.1016/j.semcdb.2013.03.013>.
- [41] R.A. Sturm, T.N. Frudakis, Eye colour: portals into pigmentation genes and ancestry, *Trends Genet.* 20 (2004) 327-332. <https://doi.org/10.1016/j.tig.2004.06.010>.
- [42] W. Oetting, R. King, Molecular basis of albinism: mutations and polymorphisms of pigmentation genes associated with albinism. *Hum. Mutat.* 13 (1999) 99-115. [https://doi.org/10.1002/\(SICI\)1098-1004\(1999\)13:2<99::AID-HUMU2>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1098-1004(1999)13:2<99::AID-HUMU2>3.0.CO;2-C).
- [43] M. Kayser, F. Liu, A.C. Janssens, F. Rivadeneira, O. Lao, K. van Duijn, M. Vermeulen, P. Arp, M.M. Jhamai, W.F. van Ijcken, J.T. den Dunnen, S. Heath, D. Zelenika, D.D. Despriet, C.C. Klaver, J.R. Vingerling, P.T. de Jong, A. Hofman, Y.S. Aulchenko, A.G. Uitterlinden, B.A. Oostra, C.M. van Duijn, Three Genome-wide Association Studies and a Linkage Analysis Identify HERC2 as a Human Iris Color Gene, *Am. J. Hum. Genet.* 82 (2008) 411-423. <https://doi.org/10.1016/j.ajhg.2007.10.003>
- [44] M. Visser, M. Kayser, R.J. Palstra, HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer

and the OCA2 promoter, *Genome Res.* 22 (2012) 446-455.
<https://doi.org/10.1101/gr.128652.111>.

[45] O. Spichenok, Z.M. Budimlija, A.A. Mitchell, A. Jenny, L. Kovacevic, D. Marjanovic, T. Caragine, M. Prinz, E. Wurmba, Prediction of eye and skin color in diverse populations using seven SNPs, *Forensic Sci. Int. Genet.* 5 (2011) 472-478.
<https://doi.org/10.1016/j.fsigen.2010.10.005>.

[46] F.R. Sawitzki, R. Rodenbusch, D.W. Gubert, M.J. Prado, D.S.B.S. Silva, E.F.A. Silva, C.S. Alho, Analysis of Eight SNPs in Pigment-Related Genes in South Brazilian Subjects with High or Low Skin and Eye Melanin Content, *SM J. Forensic Res. Criminol.* 1 (2017) 1008.

[47] Y. Ruiz, C. Phillips, A. Gomez-Tato, J. Alvarez-Dios, M.C. de Cal, R. Cruz, O. Maroñas, J. Söchtig, M. Fondevila, M.J. Rodriguez-Cid, A. Carracedo, M.V. Lareu. Further development of forensic eye color predictive tests. *Forensic Sci. Int. Genet.* 7(2013) 28-40. <https://doi.org/10.1016/j.fsigen.2012.05.009>.

[48] Donnelly M.P., P. Paschou, E. Grigorenko, D. Gurwitz, C. Barta, R.B. Lu, O.V. Zhukova, J.J. Kim, M. Siniscalco, M. New, H. Li, S.L.B. Kajuna, V.G. Manolopoulos, W.C. Speed, A.J. Pakstis, J.R. Kidd, K.K. Kidd. A global view of the OCA2-HERC2 region and pigmentation. *Hum. Genet.* 131 (2012) 683-696.
<https://doi.org/10.1007/s00439-011-1110-x>.

[49] Branicki W., U. Brudnik, A. Wojas-Pel. Interactions Between HERC2, OCA2 and MC1R May Influence Human Pigmentation Phenotype, *Annals of Hum. Genet.* 73 (2009) 160-170. <https://doi.org/10.1111/j.1469-1809.2009.00504.x>.

[50] N.G. Jablonski, G. Chaplin, Human skin pigmentation as an adaptation to UV radiation, *Proc. Natl. Acad. Sci.* 107 (2010) 8962-8968.
<https://doi.org/10.1073/pnas.0914628107>.

[51] G.S. Barsh, What Controls Variation in Human Skin Color?, *PLoS Biol.* 1 (2003) e7. <https://doi.org/10.1371/journal.pbio.0000027>.

- [52] O. Maroñas, C. Phillips, J. Söchtig, A. Gomez-Tato, R. Cruz, J. Alvarez-Dios, M.C. de Cal, Y. Ruiz, M. Fondevila, Á. Carracedo, M.V. Lareu, Development of a forensic skin colour predictive test, *Forensic Sci. Int. Genet.* 13 (2014) 34-44. <https://doi.org/10.1016/j.fsigen.2014.06.017>.
- [53] O. Spichenok, Z.M. Budimlija, A.A. Mitchell, A. Jenny, L. Kovacevic, D. Marjanovic, T. Caragine, M. Prinz, E. Wurmbach, Prediction of eye and skin color in diverse populations using seven SNPs, *Forensic Sci. Int. Genet.* 5 (2011) 472-478. <https://doi.org/10.1016/j.fsigen.2010.10.005>.
- [54] V. Mushailov, S.A. Rodriguez, Z.M. Budimlija, M. Prinz, E. Wurmbach, Assay Development and Validation of an 8-SNP Multiplex Test to Predict Eye and Skin Coloration, *J. Forensic Sci.* 60 (2015) 990-1000. <https://doi.org/10.1111/1556-4029.12758>.
- [55] S. Walsh, L. Chaitanya, K. Breslin, C. Muralidharan, A. Bronikowska, E. Pośpiech, J. Koller, L. Kovatsi, A. Wollstein, W. Branicki, F. Liu, M. Kayser, Global skin colour prediction from DNA, *Hum. Genet.* 136 (2017) 847–863. <https://doi.org/10.1007/s00439-017-1808-5>.

4 DISCUSSÃO GERAL E CONCLUSÃO

A fenotipagem forense pelo DNA está avançando cada vez mais e está a poucos passos para ser admitida na rotina de laboratórios forenses. Vários trabalhos já foram realizados estudando SNPs de genes relacionados à pigmentação de olhos e de cabelos principalmente, obtendo alta acurácia e sensibilidade na predição de olhos azuis e castanhos e cabelos ruivos. O estudo de predição de pele é mais recente e trabalhos tem demonstrado que é necessária uma maior quantidade de SNPs para predição da pele.

A população brasileira possui ancestralidade miscigenada. Não raramente, os sistemas preditores de fenótipos globais ou estrangeiros, baseados na população europeia principalmente, apresentam falhas na avaliação de amostras brasileiras devido aos altos níveis de mistura apresentados pela nossa população, nos quais a estrutura genômica pode favorecer efeitos epistáticos e pleiotrópicos não observados em populações mais homogêneas. Portanto, é importante a análise de variantes gênicas que impactem a nossa população, criando preditores específicos que consigam identificar traços fenotípicos de indivíduos locais para auxiliar os órgãos policiais em suas investigações.

Inicialmente, seguindo um trabalho prévio (Sawitzki *et al.*, 2017), no qual foram estudados modelos LMC-HMC-SW (Low Melanin Content = indivíduos com pele clara e olhos azuis, High Melanin Content = indivíduos com pele e olhos escuros, Snow White = indivíduos com pele clara e olhos escuros) para analisar a distribuição alélica de oito SNPs bialélicos, em genes relacionados à pigmentação, em indivíduos sul-brasileiros miscigenados com base nas frequências alélicas de diferentes populações humanas. Os SNPs rs12913832 e rs916977, ambos do gene *HERC2*, estão em desequilíbrio de ligação, então consideramos somente o rs12913832 nas análises para construir preditores a partir de sete SNPs (FIGURA 9 e 10).

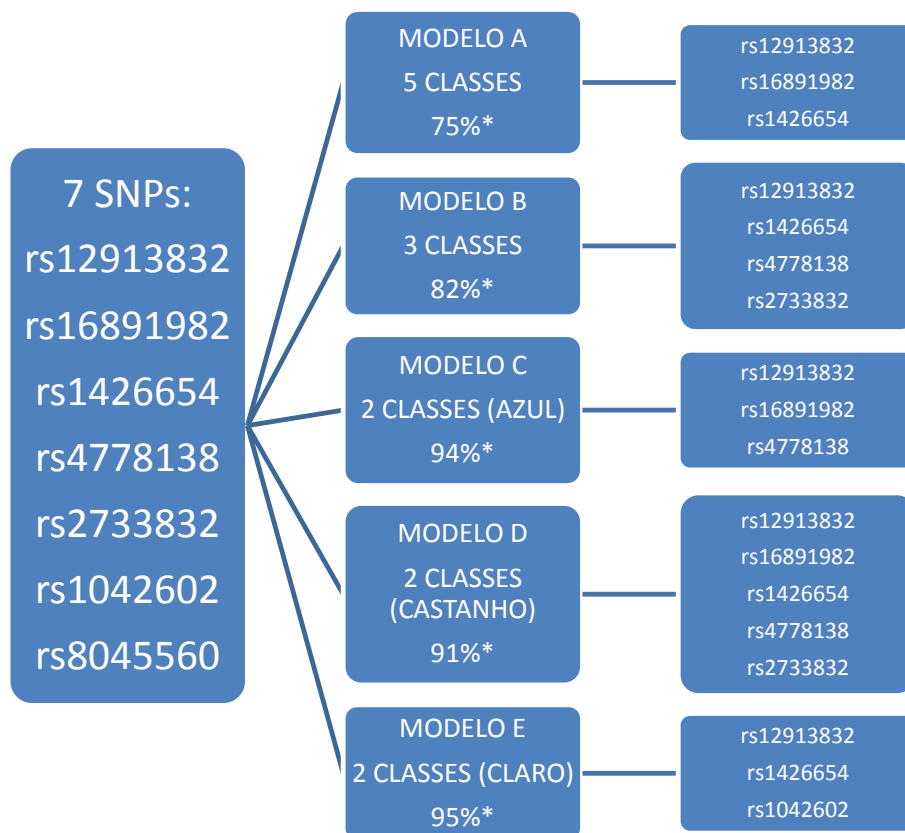


FIGURA 9: Modelos de predição de pigmentação de olhos. Os modelos utilizaram diferentes seleções de SNPs para criar os preditores. Somente o rs8045560 não encontrou significância estatística em nenhum modelo. *Acurácia.



FIGURA 10: Modelos de predição de pigmentação de pele. Os três modelos utilizaram a mesma seleção de seis SNPs para criar os preditores. Somente o rs8045560 não encontrou significância estatística em nenhum modelo. *Acurácia.

Os cinco modelos (A, B, C, D e E) de previsão de cor de olhos propostos do primeiro artigo, baseados em regressão logística multinomial, que é o método padrão ouro utilizado até hoje, são capazes de identificar fenótipos extremos (Azul ou Castanho) com alta acurácia e sensibilidade. A predição de olhos de cor intermediária ainda não é ideal, necessitando talvez do descobrimento de outros SNPs que possam segregar melhor esses fenótipos e aumentar a representatividade da cor no conjunto amostral. O modelo G de predição de cor de pele de três classes do primeiro artigo pode ser utilizado para segregar três grupos fenotípicos, obtendo uma maior acurácia e sensibilidade na previsão do fenótipo claro e do escuro e menor do fenótipo intermediário. O modelo H, de predição considerando duas classes para pele, nos mostra que o conjunto amostral poderia ser dividido em duas partes (claro e escuro) com alta acurácia, o que pode ser observado também nas análises do segundo artigo.

Após a construção dos preditores utilizando sete SNPs, foram selecionados outros SNPs relacionados à pigmentação encontrados na literatura, totalizando sessenta e oito, a fim de verificar outras variantes genéticas que possam contribuir significativamente na predição. Foram utilizados outros métodos de genotipagem e de análise estatística em relação ao artigo prévio, devido ao número maior de SNPs, que não possibilita o uso das técnicas de SNaPshot e de regressão logística multinomial. Assim, foram utilizadas técnicas de bioinformática, sendo duas técnicas de seleção de características (RFE e X^2) e três algoritmos de aprendizado de máquina (SVM, RF e MLP) para analisar os dados e construir o preditor de olhos e de pele, que foi nominado BR-EASE.

O BR-EASE é uma ferramenta de simples interface e fácil uso (FIGURA 11 e 12), que pode ser utilizada pelos órgãos policiais, centros de pesquisa na área genética e antropologia e também em sala de aula, demonstrando a atuação dos genes em conjunto (poligenia) para determinar características de fácil visualização, como o fenótipo de cor de olhos (três classes: azul, intermediário e castanho) e de pele (três classes: clara, intermediária e escura).

BR-EASE: Eye Color Estimator

Ten SNPs were selected in the following human pigmentation associated genes: ASIP (rs6119471), DDB1 (rs11230664), HERC2 (rs6497271, rs1129038, rs12913832), MFSD12 (rs10424065, rs6510760), OCA2 (rs1800407), SLC24A5 (rs1426654) and SLC24A4 (rs2402130).

SNP Panel for Eye Color Prediction

SNP	Genotype	Gene
rs6119471	GG ▼	ASIP
rs11230664	CC ▼	DDB1
rs1129038	CC ▼	HERC2
rs12913832	GG ▼	HERC2
rs6497271	GG ▼	HERC2
rs10424065	CC ▼	MFSD12
rs6510760	AA ▼	MFSD12
rs1800407	CC ▼	OCA2
rs1426654	AA ▼	SLC24A5
rs2402130	AA ▼	SLC24A4

[Click Here to Predict](#)

Version (Release): 1.0 (05/03/2019)
 Classifier: [Random Forest](#)

FIGURA 11: BR-EASE - Estimador de cor de olhos

BR-EASE: Skyn Color Estimator

Twenty Five SNPs were selected in the following human pigmentation associated genes: ASIP (rs2424984, rs6119471), DDB1 (rs11230864), HERC2 (rs1129038, rs11636232, rs16950987, rs2238289, rs4932620, rs6497271), MC1R (rs1805006, rs1805009, rs3212345, rs885479), MFSD12 (rs10424065, rs6510760), NPLOC4 (rs9894429), OCA2 (rs1037208, rs1375164, rs3794606, rs4778232), SLC24A5 (rs1426654), SLC45A2 (rs16891982, rs183671), TMEM138 (rs7948623) and TYR (rs1042602).

SNP Panel for Skin Color Prediction

SNP	Genotype	Gene
rs2424984	TT ▼	ASIP
rs6119471	GG ▼	ASIP
rs11230864	CC ▼	DDB1
rs1129038	CC ▼	HERC2
rs11636232	CC ▼	HERC2
rs16950987	AA ▼	HERC2
rs2238289	AA ▼	HERC2
rs4932620	CC ▼	HERC2
rs6497271	GG ▼	HERC2
rs1805006	CC ▼	MC1R
rs1805009	GG ▼	MC1R
rs3212345	TT ▼	MC1R
rs885479	GG ▼	MC1R
rs10424065	CC ▼	MFSD12
rs6510760	AA ▼	MFSD12
rs9894429	TT ▼	NPLOC4
rs1037208	TT ▼	OCA2
rs1375164	CC ▼	OCA2
rs3794606	AA ▼	OCA2
rs4778232	CC ▼	OCA2
rs1426654	AA ▼	SLC24A5
rs16891982	CC ▼	SLC45A2
rs183671	GG ▼	SLC45A2
rs7948623	AA ▼	TMEM138
rs1042602	CC ▼	TYR

[Click Here to Predict](#)

Version (Release): 1.0 (05/03/2019)

Classifier: [Support Vector Machine](#)

FIGURA 12: BR-EASE – Estimador de cor de pele

Os SNPs rs12913832 e rs1426654 foram coincidentes nos dois estudos considerando preditores para cor de olhos considerando três classes fenotípicas (Modelo B e BR-EASE), enquanto que os SNPs rs1426654, rs16891982 e rs1042602 foram selecionados tanto pelo Modelo G quanto pelo BR-EASE, considerando preditores para cor de pele considerando três classes fenotípicas.

Houve um ganho na predição da cor de olhos castanhos e azuis utilizando esse sistema, baseado em ferramentas de bioinformática (RFE e RF), embora a predição de olhos de cor intermediária ainda seja baixa. Além disso, a predição da cor da pele (via RFE e SVM) possui alta acurácia e sensibilidade para as três classes, gerando um resultado inédito, pois os sistemas de predição de pele costumam ter problemas na determinação de cor intermediária também.

Esses modelos de previsão de fenótipo baseados em DNA podem ser particularmente úteis em casos em que o perfil de DNA clássico não gera correspondência de banco de dados ou ajuda na classificação de cadáveres decompostos, esqueletos ou outras evidências biológicas e em casos de identificação de pessoas desaparecidas.

REFERÊNCIAS

ALALUF, S.; BARRETT, K.; BLOUNT, M.; CARTER, N. Ethnic variation in tyrosinase and TYRP1 expression in photoexposed and photoprotected human skin. **Pigment Cell Res**, v. 16, p. 35-42, 2003.

ALLWOOD, J. S.; HARBISON, S. A. SNP model development for the prediction of eye colour in New Zealand. **Forensic Sci Int Genet**, v. 7, p. 444–452, 2013.

ANCANS, J.; TOBIN, D. J.; HOOGDUIJN, M. J.; SMIT, N. P.; WAKAMATSU, K.; THODY, A. J. Melanosomal pH controls rate of melanogenesis, eumelanin/phaeomelanin ratio and melanosome maturation in melanocytes and melanoma cells. **Exp Cell Res**, v. 268, n. 1, p. 26-35, 2001.

BIESECKER, L. G.; BAILEY-WILSON, J. E.; BALLANTYNE, J.; BAUM, H.; BIEBER, F. R.; BRENNER, C.; BUDOWLE, B.; BUTLER, J. M.; CARMODY, G.; CONNEALLY, P. M.; DUCEMAN, B.; EISENBERG, A.; FORMAN, L.; KIDD, K. K.; LECLAIR, B.; NIEZGODA, S.; PARSONS, T. J.; PUGH, E.; SHALER, R.; SHERRY, S. T.; SOZER, A.; WALSH A. DNA Identifications After the 9/11 World Trade Center Attack. **Science**, v. 310, p. 1122-1123, 2005.

BAUER, C. M. *et al.* Molecular genetic investigations on Austria's patron saint Leopold III. **Forensic Sci Int Genet**, v. 7, n. 2, p. 313-5, 2013.

BERSON, J.; HARPER, D. C.; TENZA, D.; RAPOSO, G.; MARKS, M. S. Pmel17 initiates premelanosome morphogenesis within multivesicular bodies. **Molecular Biology of the cell**, v. 12, p. 3451-3464, 2001.

BILGE, Y.; KEDICI, P. S.; ALAKOÇ, Y. D.; ÜLKÜER, K. Ü.; ILKYAZ, Y. Y. The identification of a dismembered human body: a multidisciplinary approach. **Forensic Sci Int**, v. 137, p. 141-146, 2003.

BRANICKI, W.; BRUDNIK, U.; WOJAS-PEL, A. Interactions Between HERC2, OCA2 and MC1R May Influence Human Pigmentation Phenotype. **Annals of Hum Genet**, v. 73, p. 160-170, 2009.

BRENNER, M.; BERKING, M. Grundlagen der Hautpigmentierung: Biochemie und regulation der melaninsynthese. **Der Hautarzt**, v. 61, p. 554-560, 2010.

BUTLER, J. M.; SHEN, Y.; MCCORD, B. R. The development of reduced size STR amplicons as tools for analysis of degraded DNA. **J Forensic Sci**, v. 48, n. 5, p. 1054-1064, 2003.

CHAKI M.; SENGUPTA, M.; MONDAL, M.; BHATTACHARYA, A.; MALLICK, S.; BHADRA, R.; RAY K. Molecular and functional studies of tyrosinase variants among indian oculocutaneous albinism type 1. **Patients Journal of Investigative Dermatology**, v. 131, p. 260–262, 2011.

CLAES, P.; LIBERTON, D. K.; DANIELS, K.; ROSANA, K. M.; QUILLEN, E. E; PEARSON, L. N. *et al.* Modeling 3D Facial Shape from DNA. **PLoS Genet**, v. 10, p. e1004224, 2014.

COROVIC, J.; CHRISTIANSON, S. Å.; BERGMAN, L. R. From crime scene actions in stranger rape to prediction of rapist type: single-victim or serial rapist? **Behav Sci Law**, v. 30, n. 6, p. 764-781, 2012.

COSTIN, E. G., HEARING, V. J. Human skin pigmentation: melanocytes modulate skin color in response to stress. **The Journal of the Federation of American Societies for Experimental Biology**, v. 21, p. 976-994, 2007.

DIMISIANOS, G.; STEFANAKI, I.; NICOLAOU, V.; SYPSA, V.; ANTONIOU, C.; POULOU, M.; PAPADOPOULOS, O.; GOGAS, H.; KANAVAKIS, E.; NICOLAIDOU, E.; KATSAMBAS, A. D.; STRATIGOS, A. J. A study of a single variant allele (rs1426654) of the pigmentation-related gene SLC24A5 in Greek subjects. **Experimental Dermatology**, v. 18, p. 175–177, 2009.

DONNELLY, M. P.; PASCHOU, P.; GRIGORENKO, E.; GURWITZ, D.; BARTA, C.; LU, R. B.; ZHUKOVA, O. V.; KIM, J. J.; SINISCALCO, M.; NEW, M.; LI, H.; KAJUNA, S. L. B.; MANOLOPOULOS, V. G.; SPEED, W. C.; PAKSTIS, A. J.; KIDD, J. R.; KIDD, K. K.. A global view of the OCA2-HERC2 region and pigmentation. **Hum Genet**, v. 131, p. 683-696, 2012.

EIBERG, H.; TROELSEN, J.; NIELSEN, M.; MIKKELSEN, A.; MENGEL-FROM, J.; KJAER, K. W.; HANSEN, L. Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. **Hum Genet**, v. 123, n. 2, p. 177-187, 2008.

FRUDAKIS, T.; THOMAS, M.; GASKIN, Z.; VENKATESWARLU, K.; CHANDRA, K. S.; GINJUPALLI, S.; GUNTURI, S.; NATRAJAN, S.; PONNUSWAMY, V. K.; PONNUSWAMY, K. N. Sequences associated with human iris pigmentation. **Genetics**, v. 165, p. 2071–2083, 2003.

FULLER, B. B.; SPAULDING, D. T.; SMITH, D. R. Regulation of the catalytic activity of preexisting tyrosinase in black and Caucasian human melanocyte cell cultures. **Exp Cell Res**, v. 262, n. 2, p. 197-208, 2001.

GERSTENBLITH, M. R.; GOLDSTEIN, A. M.; FARGNOLI, M. C.; PERIS, K.; LANDI, M. T. Comprehensive evaluation of allele frequency differences of MC1R variants across populations. **Human Mutat**, v. 28, p. 495-505, 2007.

GIARDINA, E.; SPINELLA, A.; NOVELLI, G. Past, present and future of forensic DNA typing. **Nanomedicine**, v. 6, p. 257–270, 2011.

GIL, P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. **Int J Legal Med**, v. 114, p. 204–210, 2001.

GINGER, R. S.; ASKEW, S. E.; OGBORNE, R. M.; WILSON, R. M.; FERDINANDO, D.; DADD, T.; SMITH, A. M.; KAZI, S.; SZERENCSEI, R. T.; WINKFEIN, R. J.; SCHNETKAMP, P. P.; GREEN, M. R. SLC24A5 encodes a trans-Golgi network protein with potassium-dependent sodium–calcium exchange activity that regulates human epidermal melanogenesis. **J Biol Chem**, v. 283, p. 5486–5495, 2008.

GRAF, J.; VOISEY, J.; HUGHES, I.; VAN DAAL, A. Promoter polymorphisms in the MATP (SLC24A5) gene are associated with normal skin color variation. **Hum Mutat**, v. 28, p. 710-717, 2007.

GRAF, J.; HODGSON, R.; VAN , A. Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. **Hum Mutat**, v. 25, p. 278-284, 2005.

GRAW, M.; WEISSER, H. J.; LUTZ, S. DNA typing of human remains found in damp environments. **Forensic Sci Int**, v. 113, p. 91-95, 2000.

GRECO, G.; WAKAMATSU, K.; PANZELLA, L.; ITO, S.; NAPOLITANO, A.; D'ISCHIA, M. Isomeric cysteinyl dopas provide a (photo)degradable bulk component and a robust structural element in red human hair pheomelanin. **Pigment Cell Melanoma Res**, v. 22, p. 319–327, 2009.

HAMPIKIAN, G.; WEST, E.; AKSELROD, O. The Genetics of Innocence: Analysis of 194 U.S. DNA Exonerations. **Annu Rev Genomics Hum Genet**, v. 12, p. 97–120, 2011.

HARDING, R. M.; HEALY, E.; RAY, A. J.; ELLIS, N. S.; FLANAGAN, N.; TODD, C.; DIXON, C.; SAJANTILA, A.; JACKSON, I. J.; BIRCH-MACHIN M. A, Rees, J. L. Evidence for Variable Selective Pressures at MC1R. **Am J Human Genet**, v. 66, p. 1351-1361, 2000.

HO, H.; KAPADIA, R.; AL-TAHAN, S.; AHMAD, S.; GANESAN, A. K. WIP1 coordinates melanogenic gene transcription and melanosome formation via TORC1 inhibition. **J Biol Chem**, v. 286, n. 14, p. 12509-12523, 2011.

INMAN, K.; RUDIN, N. The origin of evidence. **Forensic Sci Int**, v. 126, n. 1, p. 11-16, 2002.

INTERNATIONAL FEDERATION OF PIGMENT CELL SOCIETIES – IFPCS 2009: Disponível em: <http://www.espcr.org/micemut/> - Acessado em abril de 2013.

ITO, S.; WAKAMATSU, K. Chemistry of Melanins. In: J.J. Nordlund, R.E. Boissy, V.J. Hearing, R.A. King, W.S. Oetting, J.P. Ortonne (Eds.). **The Pigmentary System. Physiology and Pathophysiology**, Blackwell Publishing; Oxford, p. 282-310, 2006.

ITO, S.; WAKAMATSU, K. Human hair melanins: what we have learned and have not learned from mouse coat color pigmentation. **Pigment Cell Melanoma Res**, v. 24, n. 1, p. 63-74, 2010.

JEFFREYS, A. J.; WILSON, V.; THEIN, S. L. Individual-specific 'fingerprints' of human DNA. **Nature**, v. 316, p. 76-79, 1985.

JIN, Y.; BIRLEA, S. A.; FAIN, P. R.; FERRARA, T. M.; BEN, S.; RICCARDI, S. L.; COLE, J. B.; GOWAN, K.; HOLLAND, P. J.; BENNETT, D. C.; LUITEN, R. M.; WOLKERSTORFER, A.; VAN DER VEEN, J. P. W.; HARTMANN, A.; EICHNER, S.; SCHULER, G.; VAN GEEL, N.; LAMBERT, J.; KEMP, E. H.; GAWKRODGER, D. J.; WEETMAN, A. P.; TAÏEB, A.; JOUARY, T.; EZZEDINE, K.; WALLACE, M. R.; MCCORMACK, W. T.; PICARDO, M.; LEONE, G.; OVERBECK, A.; SILVERBERG, N. B.; SPRITZ, R. A. Genomewide association analyses identify 13 new susceptibility loci for generalized vitiligo. **Nature Genetics**, v. 44, p. 676–680, 2012.

KAYSER, M.; KNIJFF, P. Improving human forensics through advances in genetics, genomics and molecular biology. **Nature Reviews**, v. 12, p. 179-192, 2011.

KAYSER, M.; SCHNEIDER, P. DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations. **Forensic Sci Int**, v. 3, p. 154–161, 2009.

KENNY, E. E.; TIMPSON, N. J.; SIKORA, M.; YEE, M. C.; MORENO-ESTRADA, A.; ENG, C.; HUNTSMAN, S.; BURCHARD, E. G.; STONEKING, M.; BUSTAMANTE, C. D.; MYLES S. Melanesian blond hair is caused by an amino acid change in TYRP1. **Science**, v. 336, p. 554, 2012.

KIDD, K. K. *et al.* Developing a SNP panel for forensic identification of individuals. **Forensic Sci Int**, v. 164, p. 20-32, 2006.

KOBACHUK, L. D. G. Estudo de frequências alélicas de dez locos STRs do cromossomo X na população do estado do Paraná e sua contribuição na identificação humana, 2012. <https://acervodigital.ufpr.br/handle/1884/27062?show=full>. Acessado em: 30/06/2017.

KONDO, T.; HEARING, V. J. Update on the regulation of mammalian melanocyte function and skin pigmentation. **Expert Rev Dermatol**, v. 6, p. 97–108, 2011.

KRAWCZAK, M. Informativity assessment for biallelic single nucleotide polymorphisms. **Electrophoresis**, v. 20, p. 1676–1681, 1999.

KUSHIMOTO, T.; BASRUR, V.; VALENCIA, J.; MATSUNAGA, J.; VIEIRA, W. D.; FERRANS, V. J.; MULLER, J.; APPELLA, E.; HEARING, V. J. A model for melanosome biogenesis based on the purification and analysis of early melanosomes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, p. 10698-10703, 2001.

LAMASON, R. L.; MOHIDEEN, M. A.; MEST, J. R.; WRONG, A. C.; NORTON, H. L.; AROS, M. C.; JURYNEC, M. J.; MAO, X.; HUMPHREVILLE, V. R.; HUMBERT, J. E.; SINHA, S.; MOORE, J. L.; JAGADEESWARAN, P.; ZHAO, W.; NING, G.; MAKALOWSKA, I.; MCKEIGUE, P. M.; O'DONNELL, D.; KITTLES, R.; PARRA, E. J.; MANGINI, N. J.; GRUNWALD, D. J.; SRIVER, M. D.; CANFIELD, V.A.; CHENG, K. C. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. **Science**, v. 310, p. 1782–1786, 2005.

LAND, E. J.; ITO, S.; WAKAMATSU, K.; RILEY P. A. Rate constants for the first two chemical steps of Eumelanogenesis. **Pigment Cell Res**, v. 16, p. 487–493, 2003.

LEBEAU, J. L. The journey to rape: Geographic distance and the rapist's method of approaching the victim. **Journal of Police Science and Administration**, v. 15, n. 2, p. 129–136, 1987.

LEITE, D. S. *et al.* Paleogenetic Studies in Guajajara Skeletal Remains, Maranhão State, Brazil. **Journal of Anthropology**, v. 2014, p. 1-8, 2014.

LINS, T. C.; VIEIRA, R.G.; ABREU, B. S.; GRATTAPAGLIA, D.; PEREIRA, R. W. Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs. **Am J Hum Biol**, v. 22, p. 187-92, 2010.

LIU, F.; WENB, B.; KAYSER, M. Colorful DNA polymorphisms in humans. **Semin Cell Dev Biol**, v. 24, n. 6-7, p. 562- 575, 2013.

LIU, F.; VAN DUIJN, K.; VINGERLING, J.R.; HOFMAN, A.; UITTERLINDEN, A. G.; JANSSENS, A. C. J. W.; KAYSER, M. Eye color and the prediction of complex phenotypes from genotypes. **Curr Biol**, v. 19, p. R192–R193, 2009.

LOWE, A.; RICHARDSON, P.; WIVELL, R.; GILL, P.; MURRAY, C.; TULLY, G.; WHITAKER, J. Use of low copy number DNA in forensic inference. **International Congress Series**, v. 1239, p. 799–801, 2003.

MAKOVA, K.; NORTON, H. Worlwide polymorphism at the MC1R locus and normal pigmentation variation in humans **Peptides**, v. 26, p. 1901-1908, 2005.

MARINHO, A. N. R. *et al.* Paleogenetic and taphonomic analysis of human bones from Moa, Beirada, and Zé Espinho Sambaquis, Rio de Janeiro, Brazil. **Mem Inst Oswaldo Cruz** [online], v.101, n. 2, p. 15-23, 2006.

NAKAYAMA, K.; SOEMANTRI, A.; BUMBEIN DASHNYAM, F. J.; OHTSUKA, R.; DUANCHANG, P.; ISA, M. N.; SETTTHEETHAM-ISHIDA, W.; HARIHARA, S.; ISHIDA, T. Identification of novel functional variants of the melanocortin 1 receptor gene originated from asians. **Human Genet**, v. 119, p. 322-30, 2006.

NAPOLITANO, A.; COSTANTINI, C.; CRESCENZI, O.; PROTA, G. Characterization of 1,4- benzothiazine intermediates in the oxidative conversion of 5-S-cysteinyl-dopa to pheomelanins. **Tetrahedron Lett**, v. 35, p. 6365–6368, 1994.

NAPOLITANO, A.; DI DONATO, P.; PROTA, G.; LAND, E. J. Transient quinonimines and 1,4-benzothiazines of pheomelanogenesis: new pulse radiolytic and spectrophotometric evidence. **Free Radic Biol Med**, v 27, p. 521–528, 1999.

NAPOLITANO, A.; DI DONATO, P.; PROTA, G. New regulatory mechanisms in the biosynthesis of pheomelanins: rearrangement vs. redox exchange reaction routes of a transient 2H-1,4-benzothiazine-oquinonimine intermediate. **Biochim Biophys Acta**, v. 1475, p. 47–54, 2000.

NORTON, H. L.; KITTLES, R. A.; PARRA, E.; MCKEIGUE, P.; MAO, X.; CHENG, K.; CANFIELD, V. A.; BRADLEY, D. G.; MCEVOY, B.; SHRIVER, M. Genetic evidence for the convergent evolution of light skin in Europeans and East Asian. **Mol Biol Evol**, v. 24, p. 710-22, 2007.

O'RAHILLY, S. Human genetics illuminates the path to metabolic disease. **Nature**, v. 462, p. 307-314, 2009.

PALUMBO, A.; SOLANO, F.; MISURACA, G.; AROCA, P.; GARCIA-BORRÓN, J. C.; LOZANO, J. A.; PROTA, G. Comparative action of dopachrome tautomerase and metal ions on the rearrangement of dopachrome. **Biochim Biophys Acta**, v. 1115, p. 1-5, 1991.

PALUMBO, A.; D'ISCHIA, M.; PROTA, G. Tyrosinase-promoted oxidation of 5,6-dihydroxyindole-2-carboxylic acid to melanin. Isolation and characterization of oligomer intermediates. **Tetrahedron**, v. 43, p. 4203-4206, 1987.

PARRA, E. J. Human pigmentation variation: Evolution, genetic basis, and implications for public health. **Year book of physical anthropology**, v. 50, p. 85-105, 2007.

PETERSON, J.; JOHNSON, D.; HERZ, D.; GRAZIANO, L.; OEHLER, T. Special Report: Sexual Assault Kit Backlog Study. U.S. Department of Justice, Office of Justice Programs, **National Institute of Justice**, 2012.

PHILLIPS, C.; FANG, R.; BALLARD, D.; FONDEVILA, M.; HARRISON, C.; HYLAND, F.; MUSGRAVE-BROWN, E.; PROFF, C.; RAMOS-LUIS, E.; SOBRINO, B.; CARRACEDO, A.; FURTADO, M. R.; SYNDERCOMBE COURT, D.; SCHNEIDER, P. M. Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel. **Forensic Sci Int: Genetics**, v. 1, p. 180-185, 2007.

REES, J. L. The genetics of Sun sensitivity in humans. **Am J Hum Genet**, v. 75, p. 739-751, 2004.

SAWITZKI, F. R.; RODENBUSCH, R.; GUBERT, D. W.; PRADO, M. J.; SILVA, D. S. B. S.; SILVA, E. F. A.; ALHO, C. S. Analysis of Eight SNPs in Pigment-Related Genes in South Brazilian Subjects with High or Low Skin and Eye Melanin Content. **SM J Forensic Res Criminol**, v. 1, n. 2, p. 1008, 2017.

SINHA, S. K.; BUDOWLE, B.; ARCOT, S. S.; RICHEY, S. L.; CHAKRABORTY, R.; JONES, M. D. *et al.* Development and validation of a multiplexed Y-chromosome STR

genotyping system, Y-PLEX 6, for forensic casework. **J Forensic Sci**, v. 48, n. 1, p. 93-103, 2003.

SOBRINO, B.; BRION, M.; CARRACEDO, A. SNPs in forensic genetics: a review on SNP typing methodologies. **Forensic Sci Int**, v. 154, p. 181–194, 2005.

SOEJIMA, M.; KODA, Y. Population differences of two coding SNPs in pigmentation-related genes SLC24A5 and SLC24A5. **Int J Legal Med**, v. 121, p. 36-39, 2007.

STOKOWSKI, R. P.; PANT, P. V. K.; DADD, T.; FEREDAY, A.; HINDS, D. A.; JARMAN, C.; FILSELL, W.; GINGER, R. S.; GREEN, M. R.; VAN DER OUDERAA, F. J.; COX, D. R. A genomewide association study of skin pigmentation in a south asian population, **The American Journal of Human Genetics**, v. 81, p. 1119–1132, 2007.

STURM, R. A.; FRUDAKIS, T. N. Eye colour: portals into pigmentation genes and ancestry. **Trends in Genetics**, v. 20, p. 327–332, 2004.

STURM, R. A.; DUFFY, D. L. Human pigmentation genes under environmental selection. **Genome Biology**, v. 13, p. 248, 2012.

STURM, R. A.; DUFFY, D. L.; ZHAO, Z. Z.; LEITE, F. P. N.; STARK, M. S.; HAYWARD, N. K.; MARTIN, N. G.; MONTGOMERY, G. W. A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. **The American Journal of Human Genetics**, v. 82, p. 424-431, 2008.

STURM, R. A.; LARSSON, M. Genetics of human iris colour and patterns. **Pigment Cell Melanoma Res**, v. 22, p. 544-562, 2009.

STURM, R. A. A golden age of human pigmentation genetics. **Trends Genet**, v. 22, p. 464-468, 2006.

SULEM, P.; GUDBJARTSSON, D. F.; STACEY, S. N.; HELGASON, A.; RAFNAR, T.; MAGNUSSON, K. P.; MANOLESCU, A.; KARASON, A.; PALSSON, A.; THORLEIFSSON, G.; JAKOBSDOTTIR, M.; STEINBERG, S.; PALSSON, S.; JONASSON, F.; SIGURGEIRSSON, B.; THORISDOTTIR, K.; RAGNARSSON, R.; BENEDIKTSDOTTIR, K. R.; ABEN, K. K.; KIEMENEY, L. A.; OLAFSSON, J. H.; GULCHER, J.; KONG, A.; THORSTEINSDOTTIR, U.; STEFANSSON, K. Genetic determinants of hair, eye and skin pigmentation in Europeans. **Nat Genet**, v. 39, p. 1443–1452, 2007.

TOMAS, C.; STANGEGAARD, M.; BØRSTING, C.; HANSEN, A. J.; MORLING, N. Typing of 48 autosomal SNPs and amelogenin with GenPlex SNP genotyping system in forensic genetics. **Forensic Sci Int: Genetics**, vol 3, p. 1-6, 2008.

TULLY, G. Genotype versus phenotype: Human pigmentation. **Forensic Sci Int Genet**, v. 1, p. 100-104, 2007.

VALENZUELA, R. K.; ITO, S.; WAKAMATSU, K.; BRILLIANT, M. H. Prediction model validation: normal human pigmentation variation. **J Forensic Res**, v. 2, p. 139, 2011.

VIERKÖTTER, A.; KRÄMER, U.; SUGIRI, D.; MORITA, A.; YAMAMOTO, A.; KANEKO, N.; MATSUI, M.; KRUTMANN, J. Development of lentigines in german and japanese women correlates with variants in the SLC24A5. **Gene Journal of Investigative Dermatology**, v. 132, p. 733–736, 2012.

VISSER, M.; KAYSER, M.; PALSTRA, R. J. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. **Genome Res**, v. 22, p. 446-455, 2012.

WAKAMATSU, K.; OHTARA, K.; ITO, S. Chemical analysis of late stages of pheomelanogenesis: conversion of dihydrobenzothiazine to a benzothiazole structure. **Pigment Cell Melanoma Res**, v. 22, p. 474–486, 2009.

WALSH, S.; LINDENBERGH, A.; ZUNIGA, S. B.; SIJEN, T.; KNIJFF, P.; KAYSER, M.; BALLANTYNE, K. N. Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. **Forensic Sci Int Genet**, v. 5, p. 464-471, 2011a.

WALSH, S.; LIU, F.; WOLLSTEIN, A.; KOVATSI, L.; RALF, A.; KOSINIAK-KAMYSZ, A.; BRANICKI, W.; KAYSER, M. The HirisPlex system for simultaneous prediction of hair and eye colour from DNA. **Forensic Sci Int Genet**, v. 7, p. 98-115, 2013.

WALSH, S.; LIU, F.; BALLANTYNE, K. N.; VAN OVEN, M.; LAO, O.; KAYSER, M. IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. **Forensic Sci Int Genet**, v. 5, n. 3, p. 170-180, 2011b.

WICTUM, E.; KUN, T.; LINDQUIST, C.; MALVICK, J.; VANKAN, D.; SACKS, B. Developmental validation of DogFiler, a novel multiplex for canine DNA profiling in forensic casework, **Forensic Sci Int Genetics**, v. 7, p. 82–91, 2013.

WILKERSON, C. L.; SYED, N. A.; FISHER, M. R.; ROBINSON, N. L.; WALLOW, H. L.; ALBERT, D. M. Melanocytes and iris color. Light microscopic findings. **Arch Ophthalmol**, v. 114, n. 4, p. 437–442, 1996.

WOODHAMS, J.; LABUSCHAGNE, G. South African serial rapists: the offenders, their victims, and their offenses. **Sex Abuse**, v. 24, n. 6, p. 544-74, 2012.

YOSHIDA, K.; YAYAMA, K.; HATANAKA, A.; TAMAKI, K. Efficacy of extended kinship analyses utilizing commercial STR kit in establishing personal identification. **Leg Med**, v. 13, p. 12-15, 2011.

ZUPANIČ PAJNIČ, I. Genetic identification of Second World War victim's skeletal remains. Saarbrücken: **LAP LAMBERT Academic Publishing**, ISBN 978-3-659-45306-9, 2013.

ANEXO I – MATERIAIS E MÉTODOS

A metodologia está descrita nos artigos. As genotipagens no sistema SNaPshot foram realizadas no Laboratório de Genética Humana e Molecular da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS) [1]. As genotipagens por espectrometria de massas foram realizadas no Liggins Institute da Universidade de Auckland na Nova Zelândia [2-3].

[1] Sawitzki F.R., Rodenbusch R., Gubert D.W., Prado M.J., Silva D.S.B.S., Silva E.F.A., Alho C.S. Analysis of Eight SNPs in Pigment-Related Genes in South Brazilian Subjects with High or Low Skin and Eye Melanin Content. *SM J. Forensic Res. Criminol.* 2017; 1(2): 1008.

[2] <https://agenacx.com/wp-content/uploads/2017/12/iPLEX-Pro-Reagents-User-Guide-1.pdf>

[3] Stacey Gabriel, Liuda Ziaugra, and Diana Tabbaa. SNP Genotyping Using the Sequenom MassARRAY iPLEX Platform. *Current Protocols in Human Genetics*. 2009, 60(1): 2.12.1-2.12.18. <https://doi.org/10.1002/0471142905.hg0212s60>.

ANEXO II

SUPPLEMENTARY MATERIAL

BR-EASE: BRAZILIAN EYE AND SKIN ESTIMATOR. A FORENSIC WEBTOOL FOR DNA PHENOTYPING.

FIGURE S1 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – EYE – 3.1 CLASSES: Blue; Intermediate (Green / Hazel); and Brown (Light / Dark Brown)	129
FIGURE S2 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – EYE – 3.2 CLASSES: Blue; Intermediate (Green / Hazel / Light Brown); Dark Brown	129
FIGURE S3 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – EYE – 3.3 CLASSES: Light (Blue / Green); Intermediate (Hazel / Light Brown); Dark Brown.....	130
FIGURE S4 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – EYE – 4 CLASSES: Blue; Green; Hazel; Brown (Light and Dark)	130
FIGURE S5 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – EYE – 5 CLASSES: Blue; Green; Hazel; Light Brown; Dark Brown.....	131
FIGURE S6 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – SKIN – 2 CLASSES: Light (White / Pale / Beige); Dark (Light Brown / Medium Brown / Dark Brown).....	131
FIGURE S7 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – SKIN – 3 CLASSES: Light (White / Pale); Intermediate (Beige / Light Brown); Dark (Medium Brown / Dark Brown).....	132
FIGURE S8 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – SKIN – 4 CLASSES: Light (White / Pale); Beige; Light Brown; Dark (Medium Brown / Dark Brown)	132

FIGURE S9 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – SKIN – 5.1 CLASSES: Light (White / Pale); Beige; Light Brown; Medium Brown; Dark Brown.....	133
FIGURE S10 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – SKIN – 5.2 CLASSES: White; Pale; Beige; Light Brown; Dark (Medium Brown / Dark Brown).....	133
FIGURE S11 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – SKIN – 6 CLASSES: White; Pale; Beige; Light Brown; Medium Brown; Dark Brown.....	134
FIGURE S12 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY X^2 AND SVM / RF / MLP – EYE – 3.1 CLASSES: Blue; Intermediate (Green / Hazel); Brown (Light and Dark).....	135
FIGURE S13 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY X^2 AND SVM / RF / MLP – EYE – 3.2 CLASSES: Blue; Intermediate (Green / Hazel / Light Brown); Dark Brown	135
FIGURE S14 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY X^2 AND SVM / RF / MLP – EYE – 3.3 CLASSES: Light (Blue / Green); Intermediate (Hazel / Light Brown); Dark Brown.....	136
FIGURE S15 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY X^2 AND SVM / RF / MLP – EYE – 4 CLASSES: Blue; Green; Hazel; Brown (Light and Dark).....	136
FIGURE S16 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY X^2 AND SVM / RF / MLP – EYE – 5 CLASSES: Blue; Green; Hazel; Light Brown; Dark Brown.....	137
FIGURE S17 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY X^2 AND SVM / RF / MLP – SKIN – 2 CLASSES: Light (White / Pale / Beige); Dark (Light Brown / Medium Brown / Dark Brown).....	137
FIGURE S18 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY X^2 AND SVM / RF / MLP – SKIN – 3 CLASSES: Light (White / Pale); Intermediate (Beige / Light Brown); Dark (Medium Brown / Dark Brown).....	138

FIGURE S19 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY X^2 AND SVM / RF / MLP – SKIN – 4 CLASSES: Light (White / Pale); Beige; Light Brown; Dark (Medium Brown / Dark Brown)	138
FIGURE S20 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY X^2 AND SVM / RF / MLP – SKIN – 5.1 CLASSES: Light (White / Pale); Beige; Light Brown; Medium Brown; Dark Brown.....	139
FIGURE S21 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY X^2 AND SVM / RF / MLP – SKIN – 5.2 CLASSES: White; Pale; Beige; Light Brown; Dark (Medium Brown / Dark Brown).....	139
FIGURE S22 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY X^2 AND SVM / RF / MLP – SKIN – 6 CLASSES: White; Pale; Beige; Light Brown; Medium Brown; Dark Brown.....	140
FIGURE S23A – SNPs individual contribution in three eye phenotypic classes (Blue, Intermediate and Brown), in descending importance order.....	141
FIGURE S23B – SNPs individual contribution in three eye phenotypic classes (Blue, Intermediate and Brown), in descending importance order.....	142
FIGURE S24A – SNPs individual contribution in three skin phenotypic classes (Light, Intermediate and Dark), in descending importance order.....	143
FIGURE S24B – SNPs individual contribution in three skin phenotypic classes (Light, Intermediate and Dark), in descending importance order.....	144
FIGURE S24C – SNPs individual contribution in three skin phenotypic classes (Light, Intermediate and Dark), in descending importance order.....	145
FIGURE S24D – SNPs individual contribution in three skin phenotypic classes (Light, Intermediate and Dark), in descending importance order.....	146

FIGURE S24E – SNPs individual contribution in three skin phenotypic classes (Light, Intermediate and Dark), in descending importance order.....	147
TABLE S1: <u>RFE SNPS SELECTION – EYE</u> – Machine learning models (SVM, RF and MLP) average accuracy (avg) and standard deviation (std) of ten-fold cross validation.....	148
TABLE S2: <u>X² SNPS SELECTION – EYE</u> – Machine learning models (SVM, RF and MLP) average accuracy (avg) and standard deviation (std) of ten-fold cross validation.....	149
TABLE S3: <u>RFE SNPS SELECTION – SKIN 234</u> – Machine learning models (SVM, RF and MLP) average accuracy (avg) and standard deviation (std) of ten-fold cross validation.....	150
TABLE S4: <u>RFE SNPS SELECTION – SKIN 56</u> – Machine learning models (SVM, RF and MLP) average accuracy (avg) and standard deviation (std) of ten-fold cross validation.....	151
TABLE S5: <u>X² SNPS SELECTION – SKIN 234</u> – Machine learning models (SVM, RF and MLP) average accuracy (avg) and standard deviation (std) of ten-fold cross validation.....	152
TABLE S6: <u>X² SNPS SELECTION – SKIN 56</u> – Machine learning models (SVM, RF and MLP) average accuracy (avg) and standard deviation (std) of ten-fold cross validation.....	153
TABLE S7: SNP SELECTION BY RFE AND X ² – EYE 3.1 CLASSES.....	154
TABLE S8: SNP SELECTION BY RFE AND X ² – EYE 3.2 CLASSES.....	164
TABLE S9: SNP SELECTION BY RFE AND X ² – EYE 3.3 CLASSES.....	174
TABLE S10: SNP SELECTION BY RFE AND X ² – EYE 4 CLASSES.....	184
TABLE S11: SNP SELECTION BY RFE AND X ² – EYE 5 CLASSES.....	195

TABLE S12: SNP SELECTION BY RFE AND X^2 – SKIN 2 CLASSES.....	205
TABLE S13: SNP SELECTION BY RFE AND X^2 – SKIN 3 CLASSES.....	215
TABLE S14: SNP SELECTION BY RFE AND X^2 – SKIN 4 CLASSES.....	226
TABLE S15: SNP SELECTION BY RFE AND X^2 – SKIN 5.1 CLASSES	236
TABLE S16: SNP SELECTION BY RFE AND X^2 – SKIN 5.2 CLASSES	246
TABLE S17: SNP SELECTION BY RFE AND X^2 – SKIN 6 CLASSES.....	257

SNPS SELECTION BY RECURSIVE FEATURE ELIMINATION (RFE) AND PHENOTYPE CLASSIFIERS (ACCURACY) THROUGH MACHINE LEARNING METHODS: SUPPORT VECTOR MACHINE (SVM), RANDOM FOREST (RF) AND MULTILAYER PERCEPTRON (MLP)

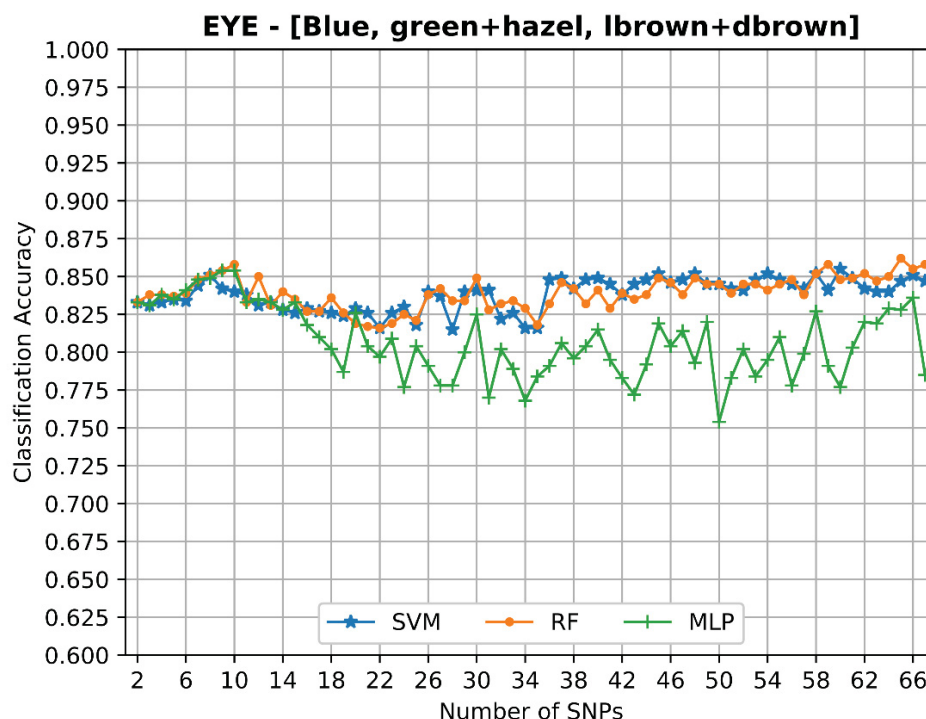


FIGURE S1 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – EYE – 3.1 CLASSES: Blue: Intermediate (Green / Hazel): and Brown (Light / Dark Brown).

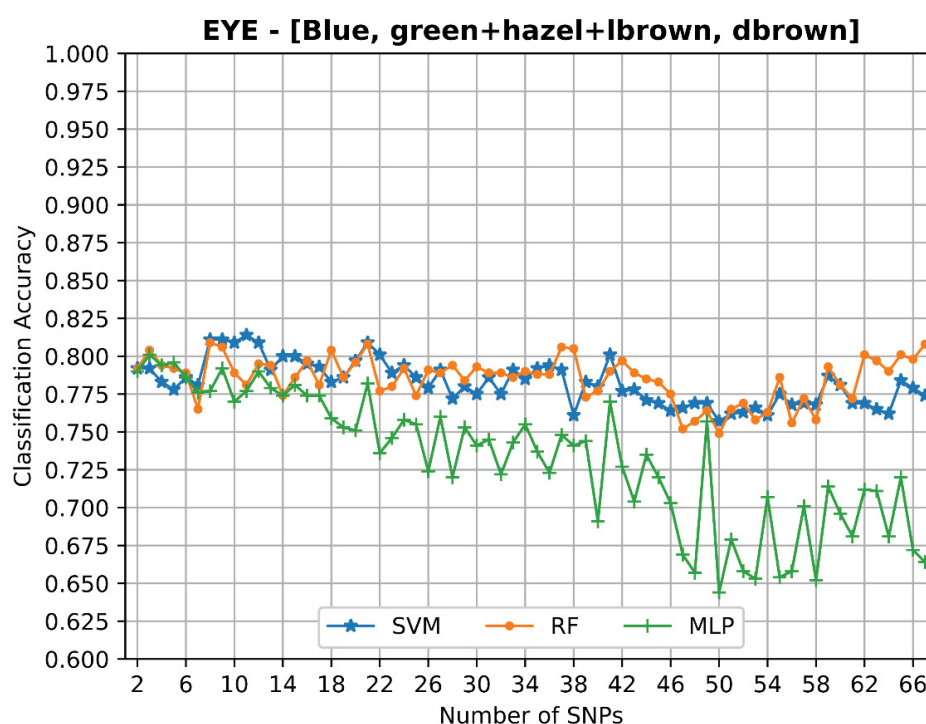


FIGURE S2 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – EYE – 3.2 CLASSES: Blue: Intermediate (Green / Hazel / Light Brown): Dark Brown.

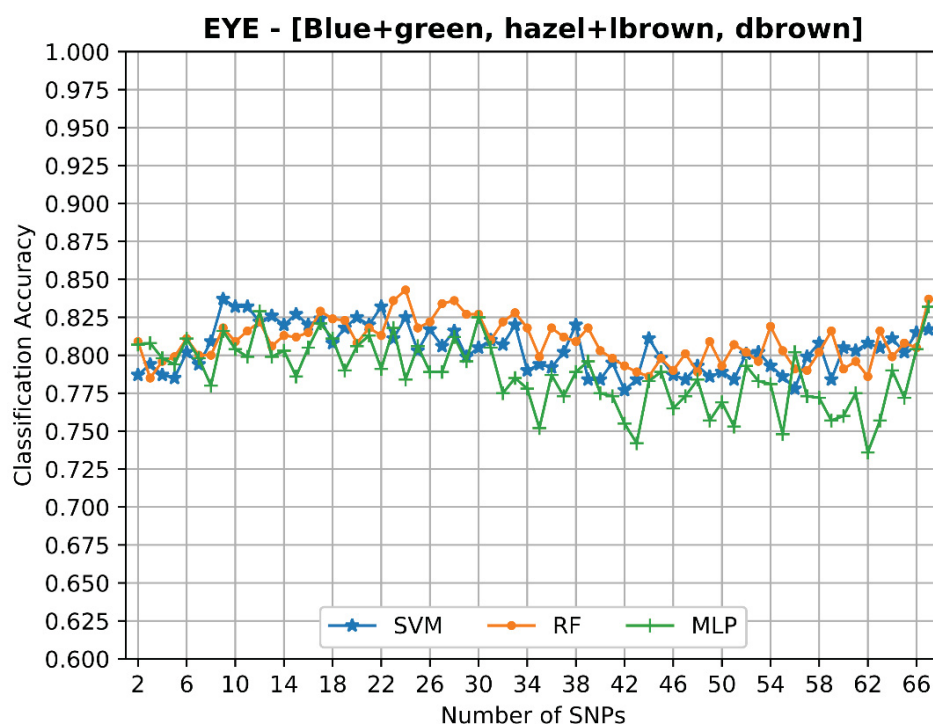


FIGURE S3 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – EYE – 3.3 CLASSES: Light (Blue / Green); Intermediate (Hazel / Light Brown); Dark Brown.

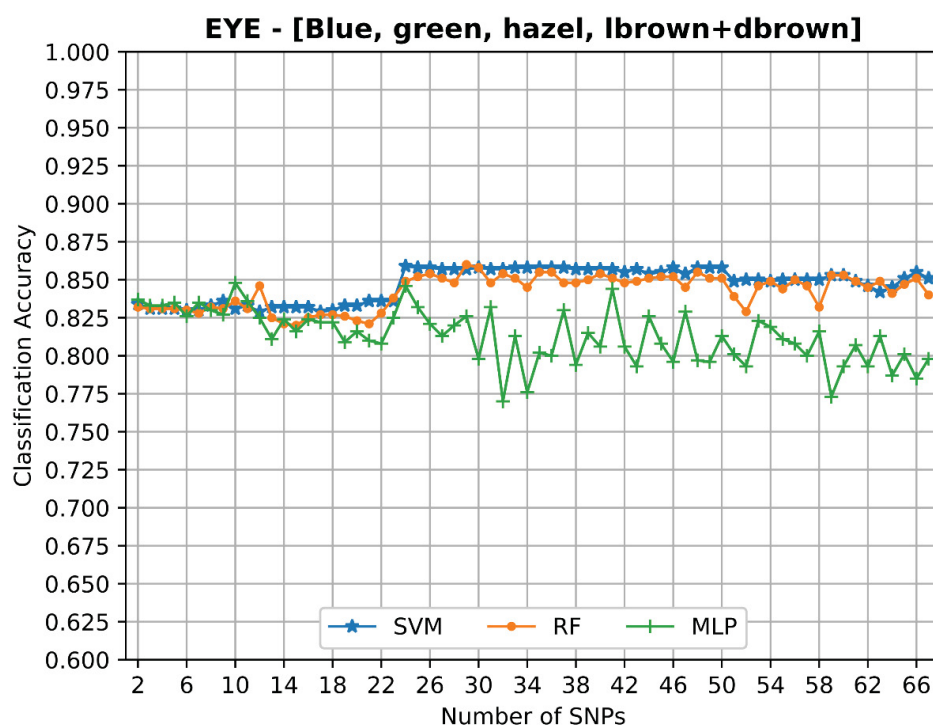


FIGURE S4 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – EYE – 4 CLASSES: Blue; Green; Hazel; Brown (Light and Dark).

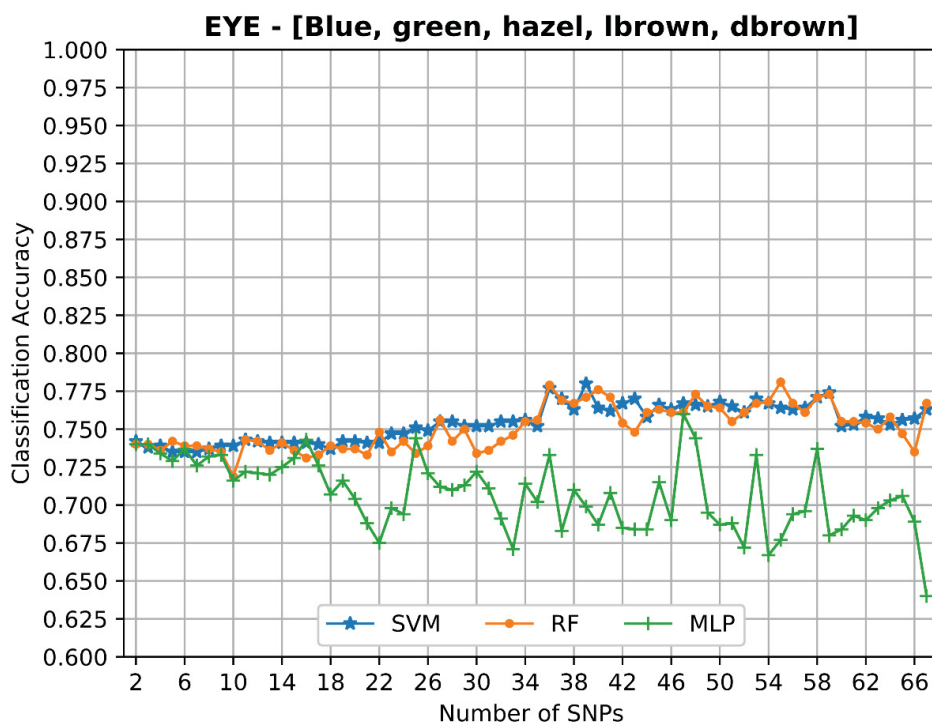


FIGURE S5 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – EYE – 5 CLASSES: Blue; Green; Hazel; Light Brown; Dark Brown.

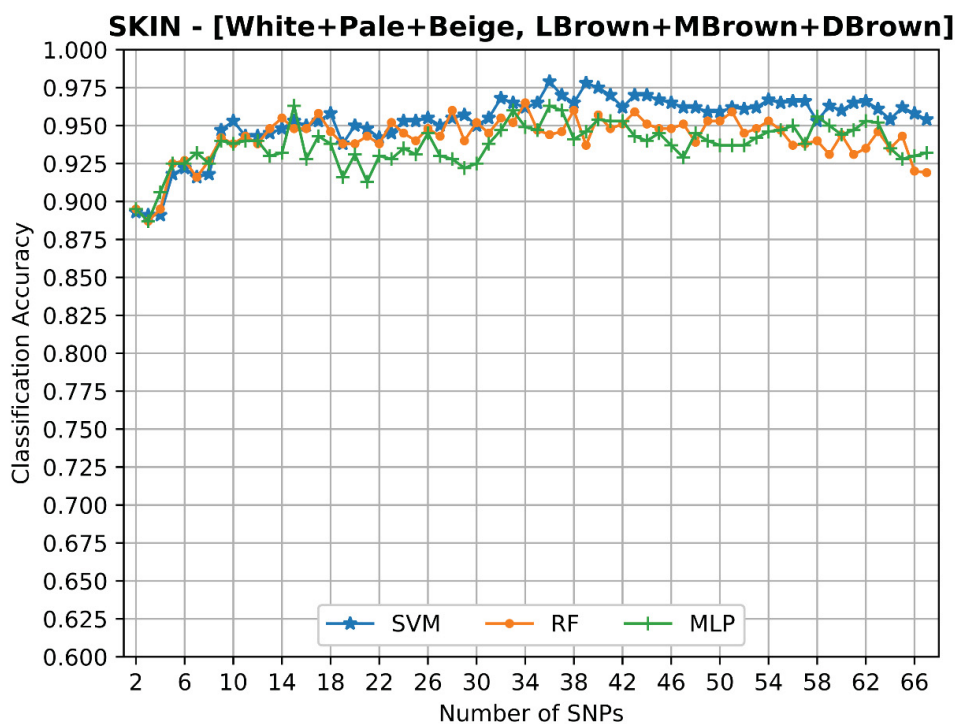


FIGURE S6 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – SKIN – 2 CLASSES: Light (White / Pale / Beige); Dark (Light Brown / Medium Brown / Dark Brown).

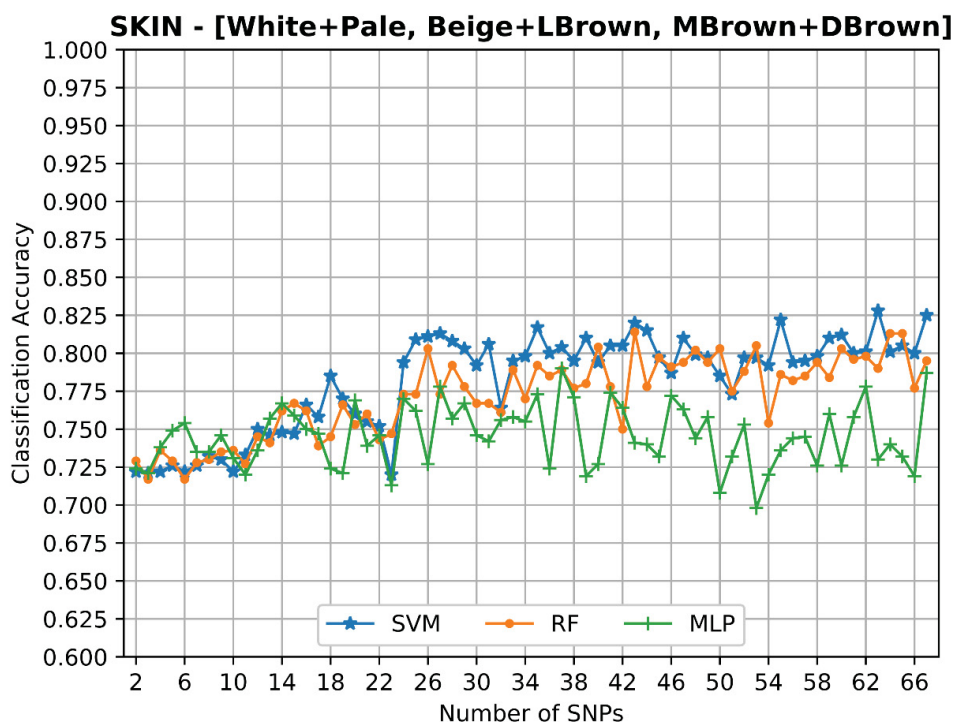


FIGURE S7 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – SKIN – 3 CLASSES: Light (White / Pale); Intermediate (Beige / Light Brown); Dark (Medium Brown / Dark Brown).

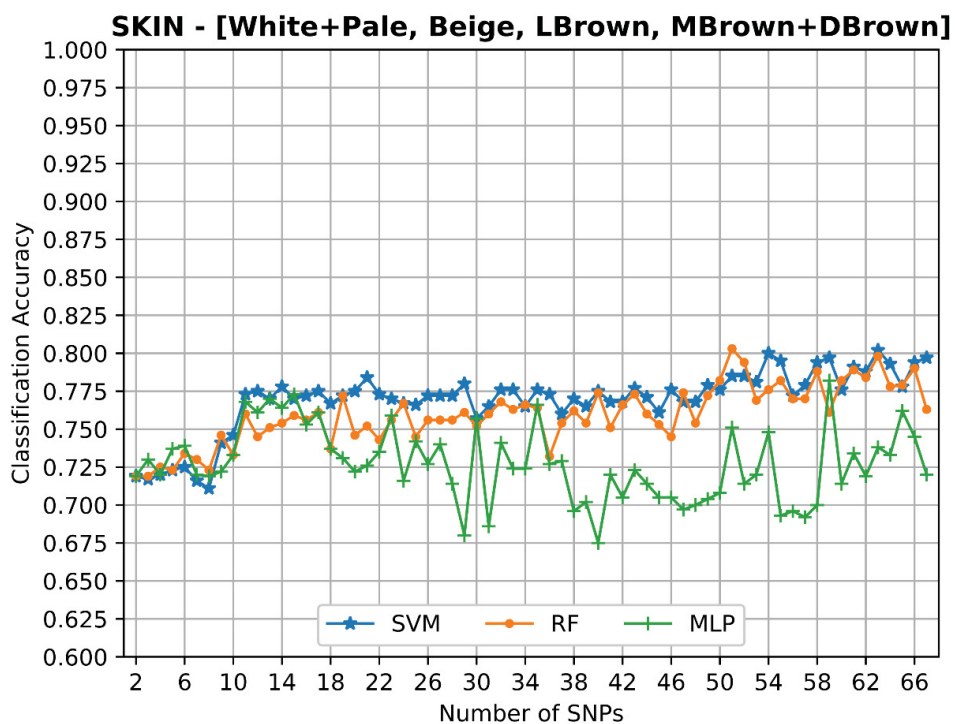


FIGURE S8 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – SKIN – 4 CLASSES: Light (White / Pale); Beige; Light Brown; Dark (Medium Brown / Dark Brown).

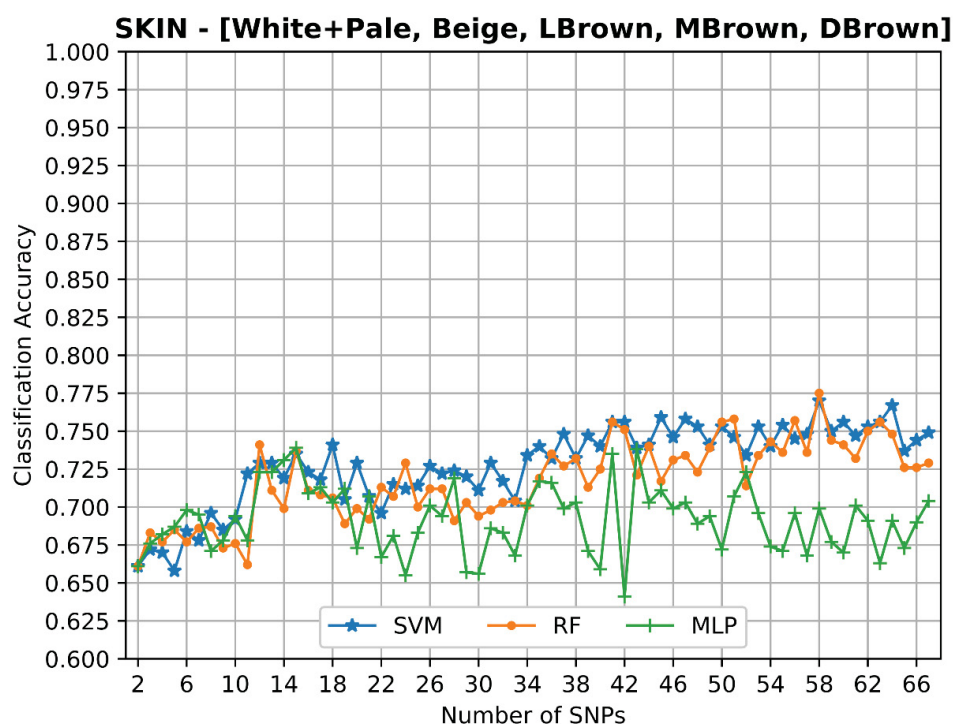


FIGURE S9 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – SKIN – 5.1 CLASSES: Light (White / Pale); Beige; Light Brown; Medium Brown; Dark Brown.

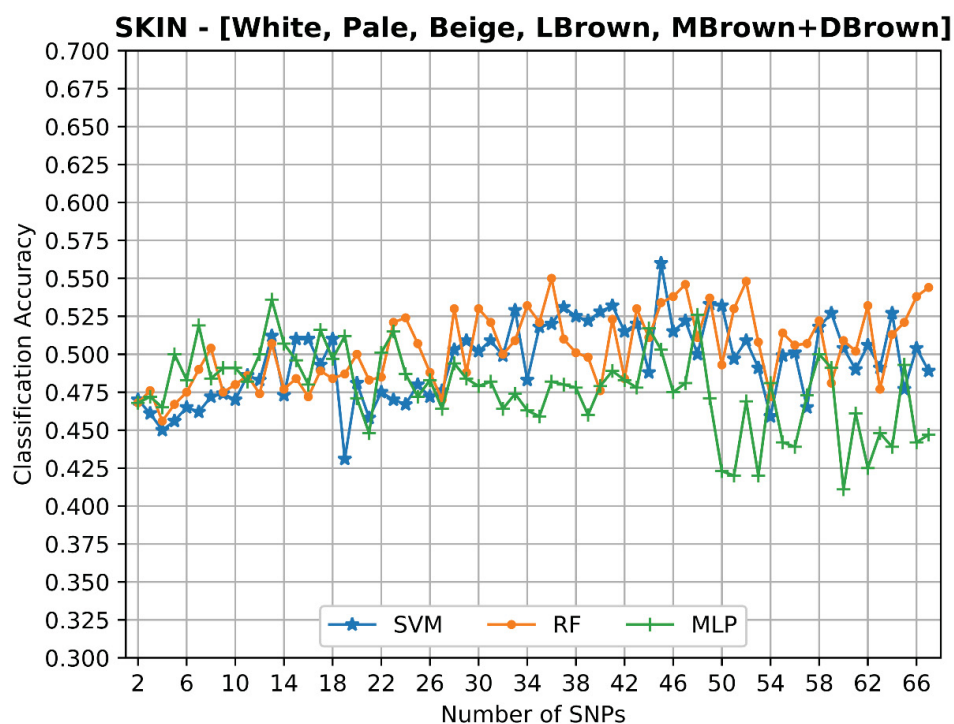


FIGURE S10 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – SKIN – 5.2 CLASSES: White; Pale; Beige; Light Brown; Dark (Medium Brown / Dark Brown).

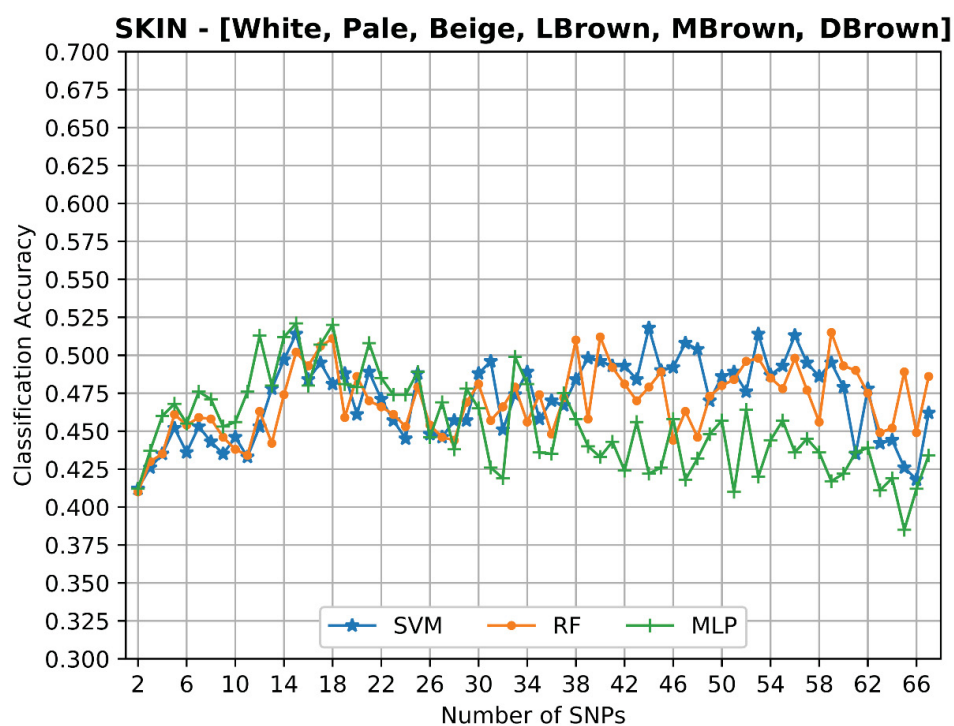


FIGURE S11 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY RFE AND SVM / RF / MLP – SKIN – 6 CLASSES: White; Pale; Beige; Light Brown; Medium Brown; Dark Brown.

SNPS SELECTION BY CHI SQUARE (χ^2) AND PHENOTYPE CLASSIFIERS (ACCURACY) THROUGH MACHINE LEARNING METHODS: SUPPORT VECTOR MACHINE (SVM), RANDOM FOREST (RF) AND MULTILAYER PERCEPTRON (MLP)

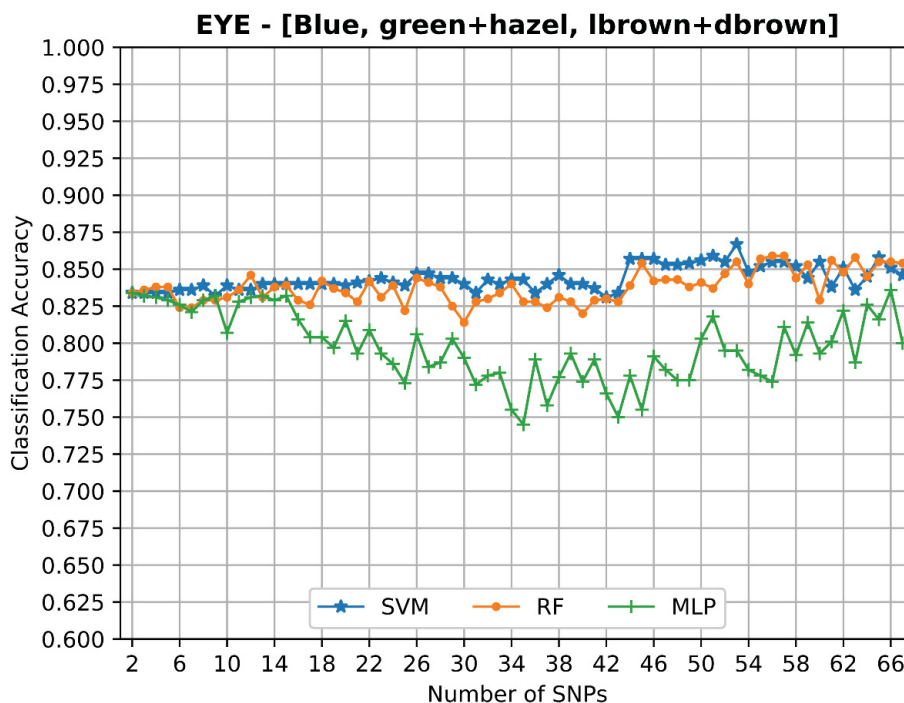


FIGURE S12 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY χ^2 AND SVM / RF / MLP – EYE – 3.1 CLASSES: Blue; Intermediate (Green / Hazel); Brown (Light and Dark).

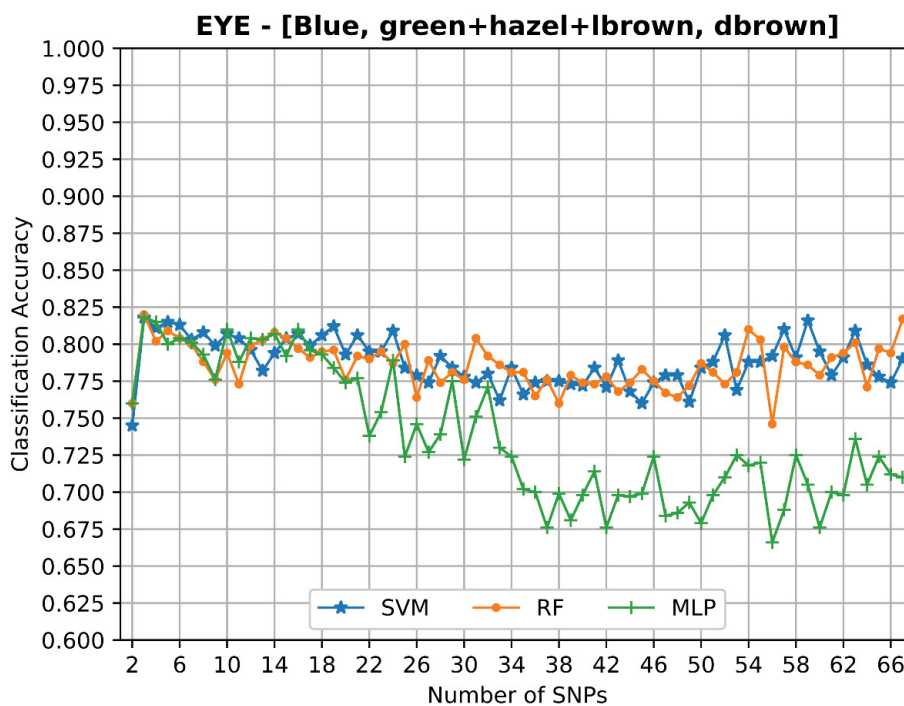


FIGURE S13 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY χ^2 AND SVM / RF / MLP – EYE – 3.2 CLASSES: Blue; Intermediate (Green / Hazel / Light Brown); Dark Brown.

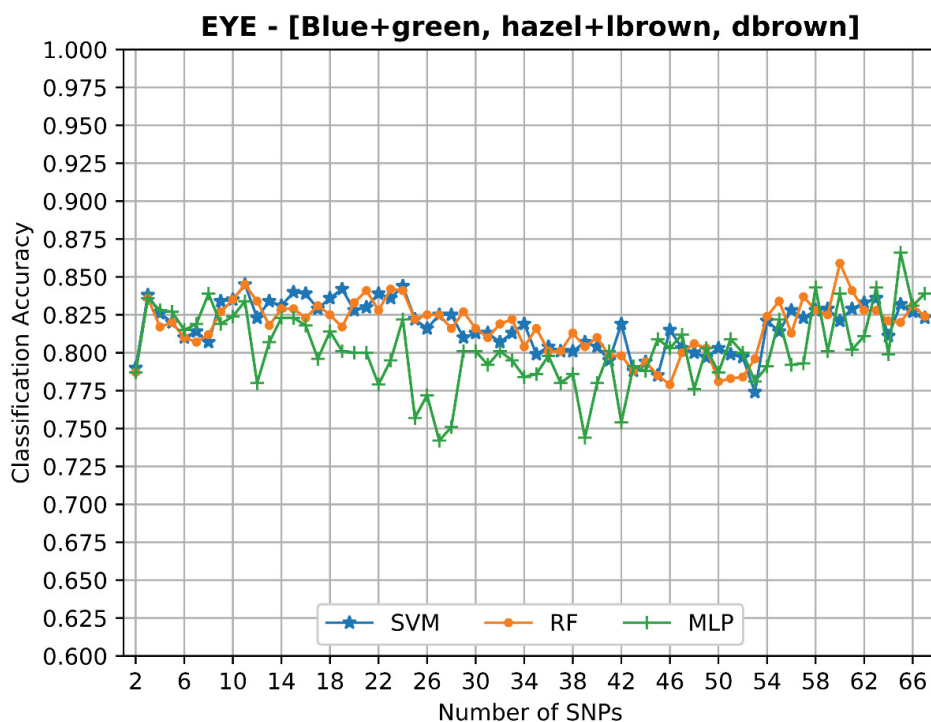


FIGURE S14 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY χ^2 AND SVM / RF / MLP – EYE – 3.3 CLASSES: Light (Blue / Green); Intermediate (Hazel / Light Brown); Dark Brown.

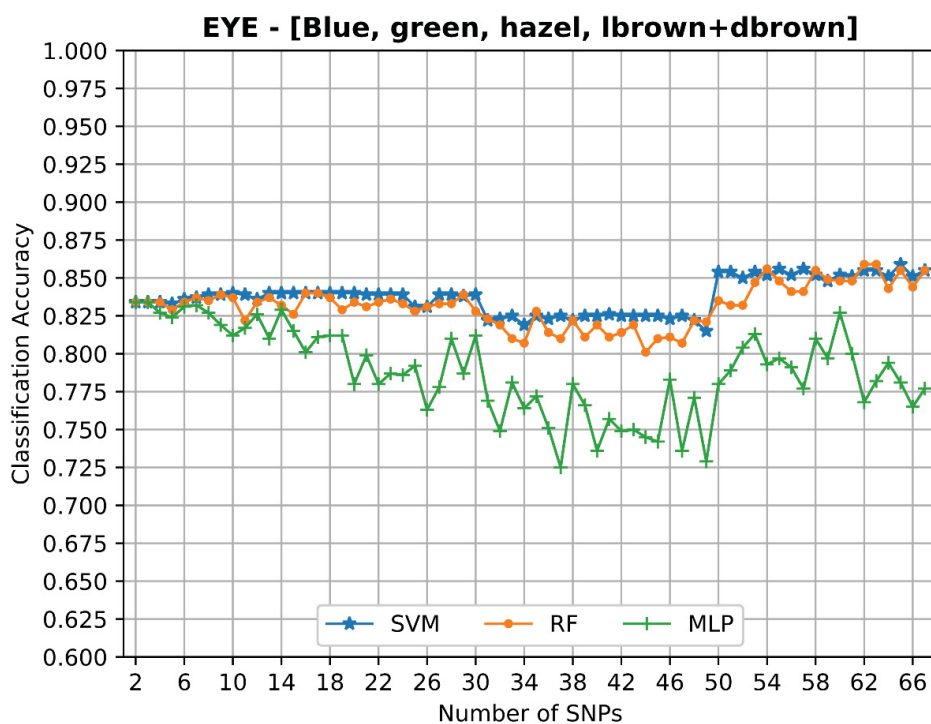


FIGURE S15 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY χ^2 AND SVM / RF / MLP – EYE – 4 CLASSES: Blue; Green; Hazel; Brown (Light and Dark).

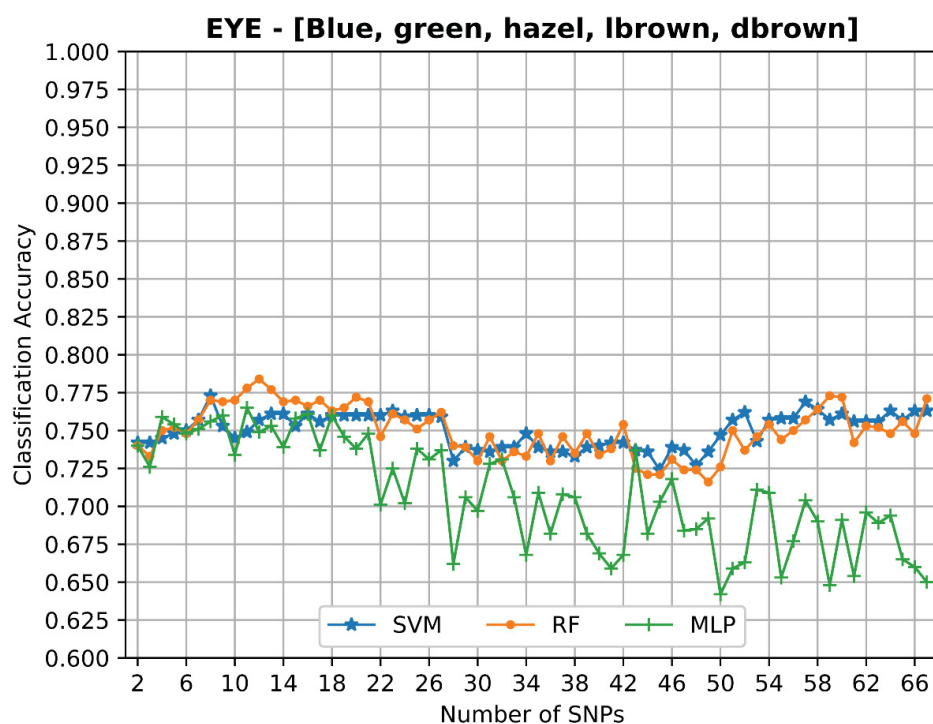


FIGURE S16 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY X^2 AND SVM / RF / MLP – EYE – 5 CLASSES: Blue; Green; Hazel; Light Brown; Dark Brown.

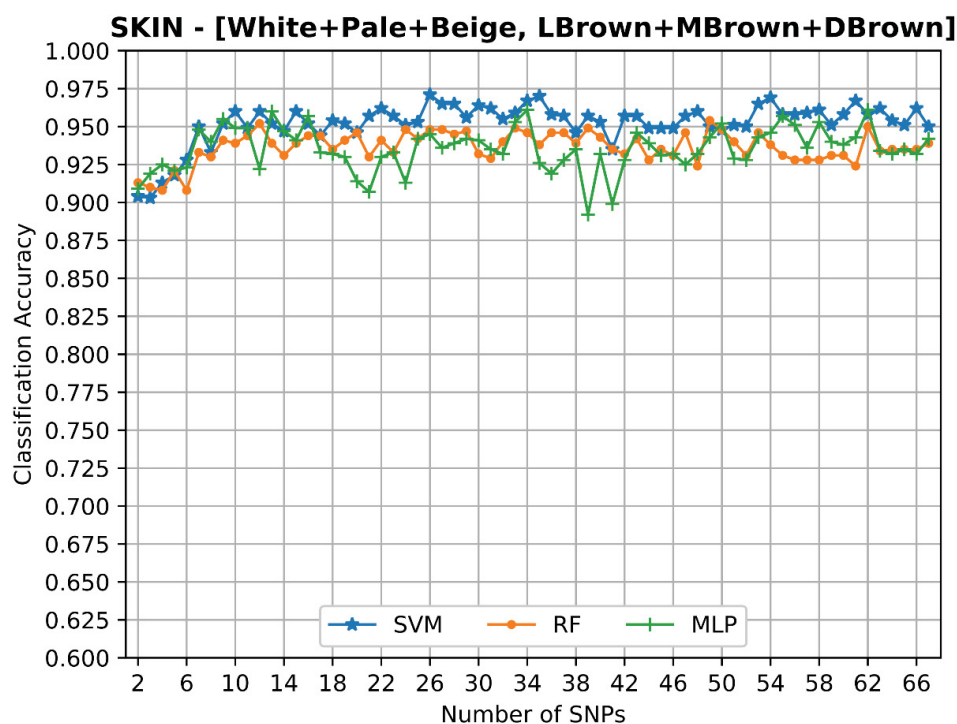


FIGURE S17 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY X^2 AND SVM / RF / MLP – SKIN – 2 CLASSES: Light (White / Pale / Beige); Dark (Light Brown / Medium Brown / Dark Brown).

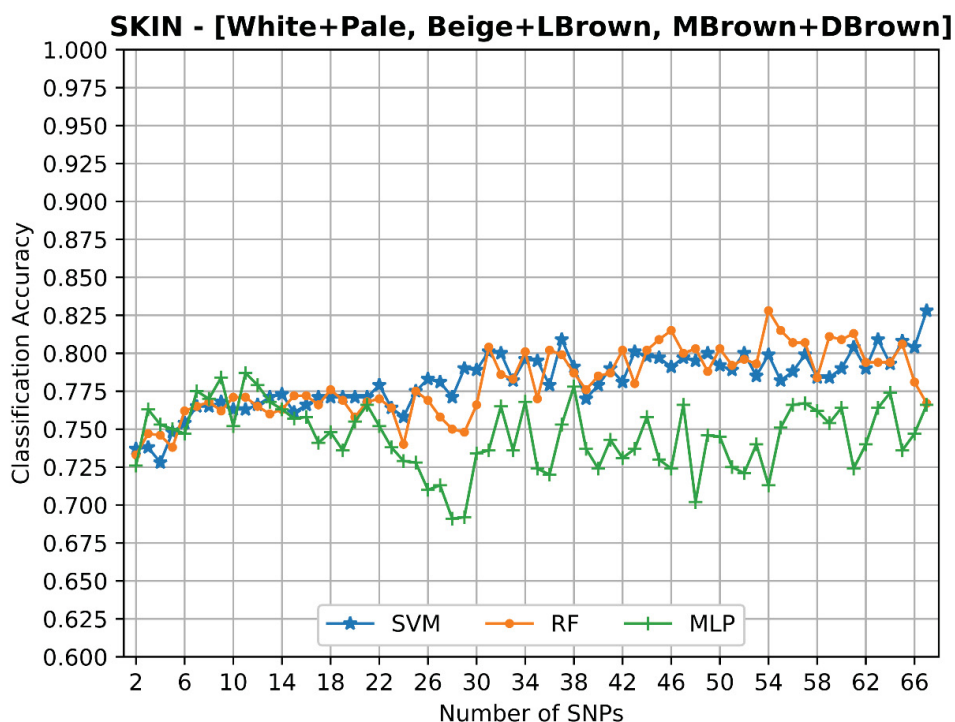


FIGURE S18 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY χ^2 AND SVM / RF / MLP – SKIN – 3 CLASSES: Light (White / Pale); Intermediate (Beige / Light Brown); Dark (Medium Brown / Dark Brown).

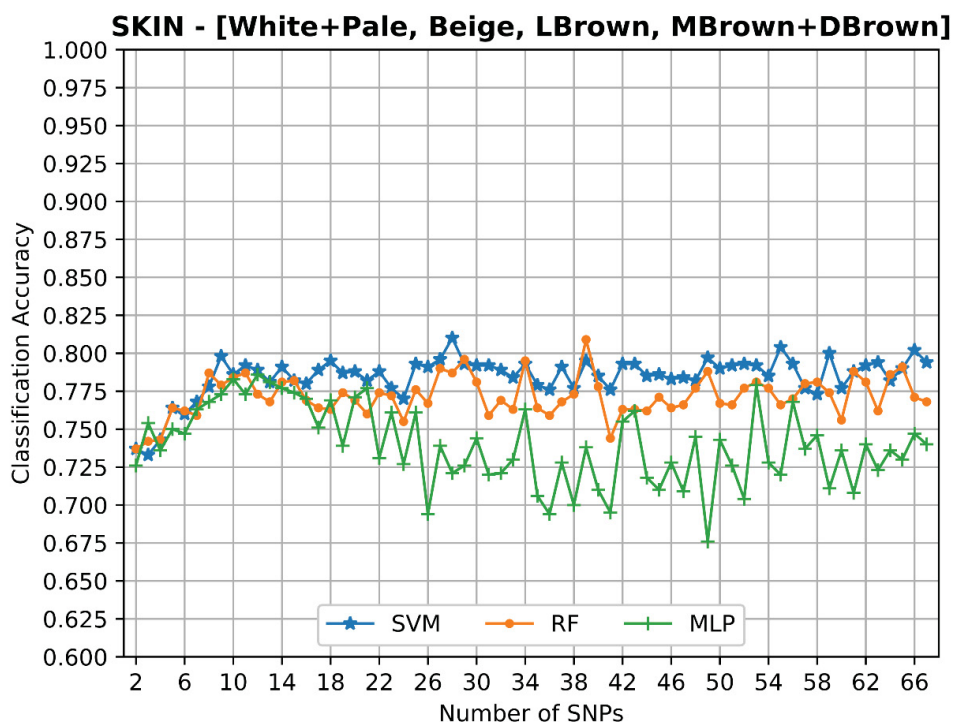


FIGURE S19 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY χ^2 AND SVM / RF / MLP – SKIN – 4 CLASSES: Light (White / Pale); Beige; Light Brown; Dark (Medium Brown / Dark Brown).

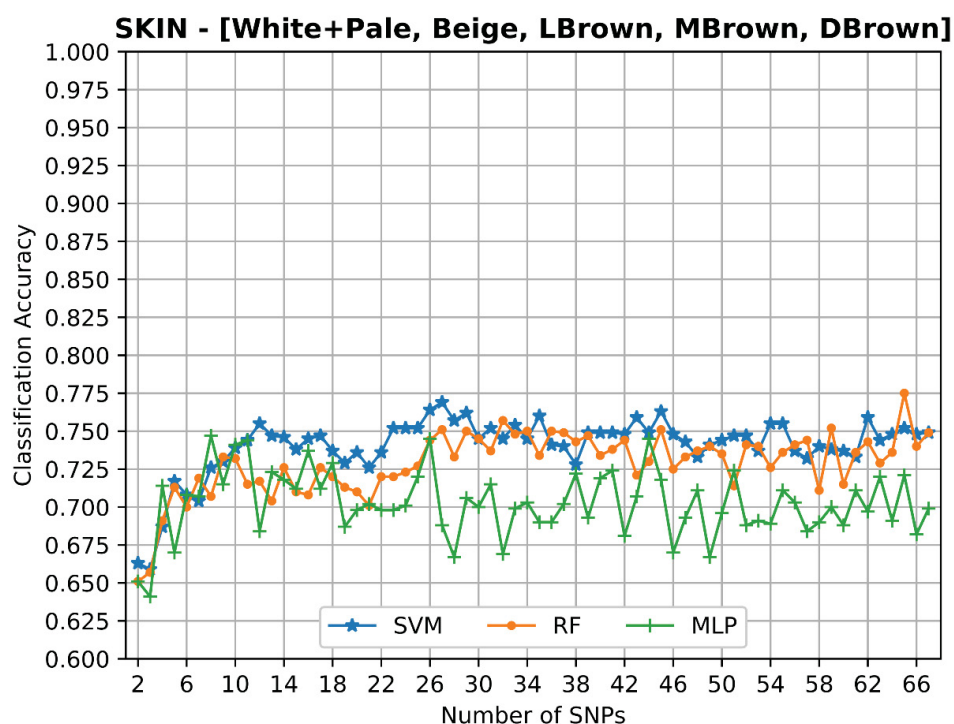


FIGURE S20 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY χ^2 AND SVM / RF / MLP – SKIN – 5.1 CLASSES: Light (White / Pale); Beige; Light Brown; Medium Brown; Dark Brown.

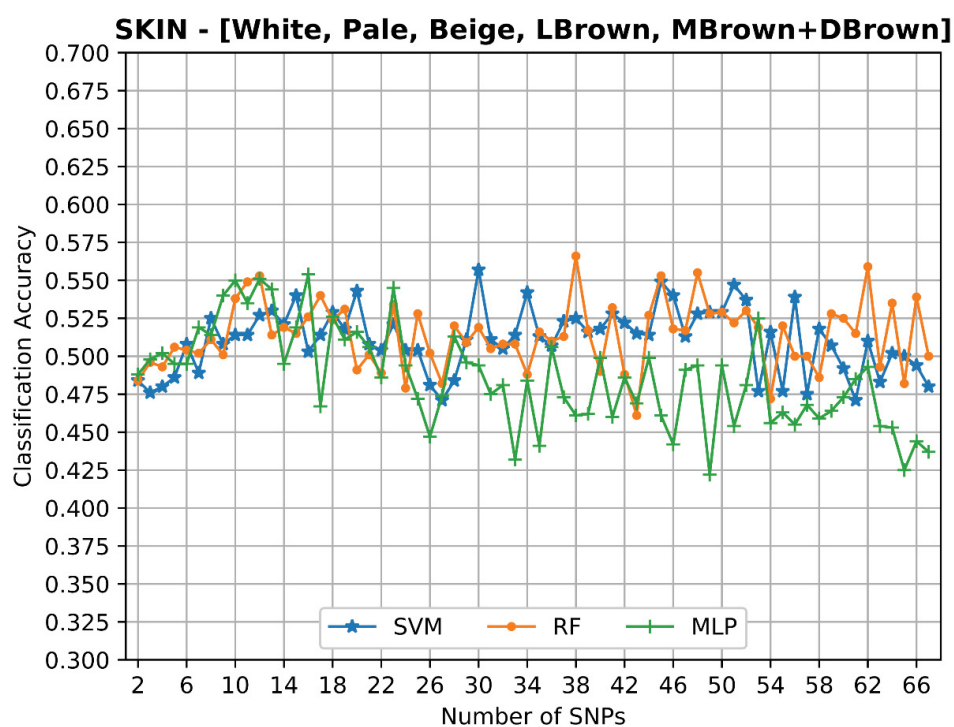


FIGURE S21 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY χ^2 AND SVM / RF / MLP – SKIN – 5.2 CLASSES: White; Pale; Beige; Light Brown; Dark (Medium Brown / Dark Brown).

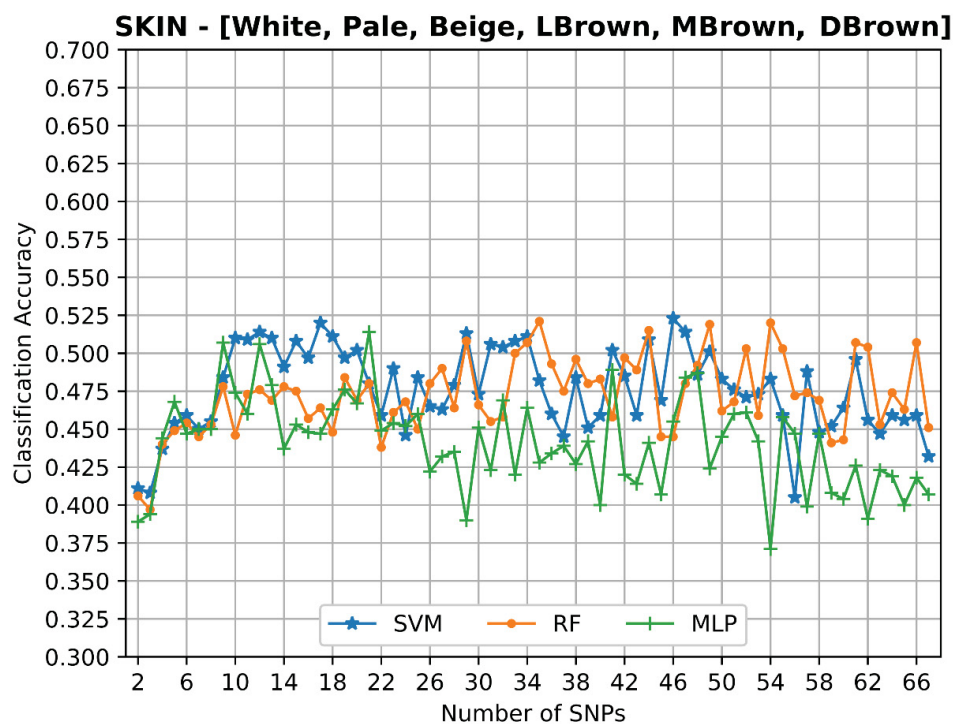


FIGURE S22 – CLASSIFICATION ACCURACY IN ACCORDANCE WITH SNPS SELECTION BY χ^2 AND SVM / RF / MLP – SKIN – 6 CLASSES: White; Pale; Beige; Light Brown; Medium Brown; Dark Brown.

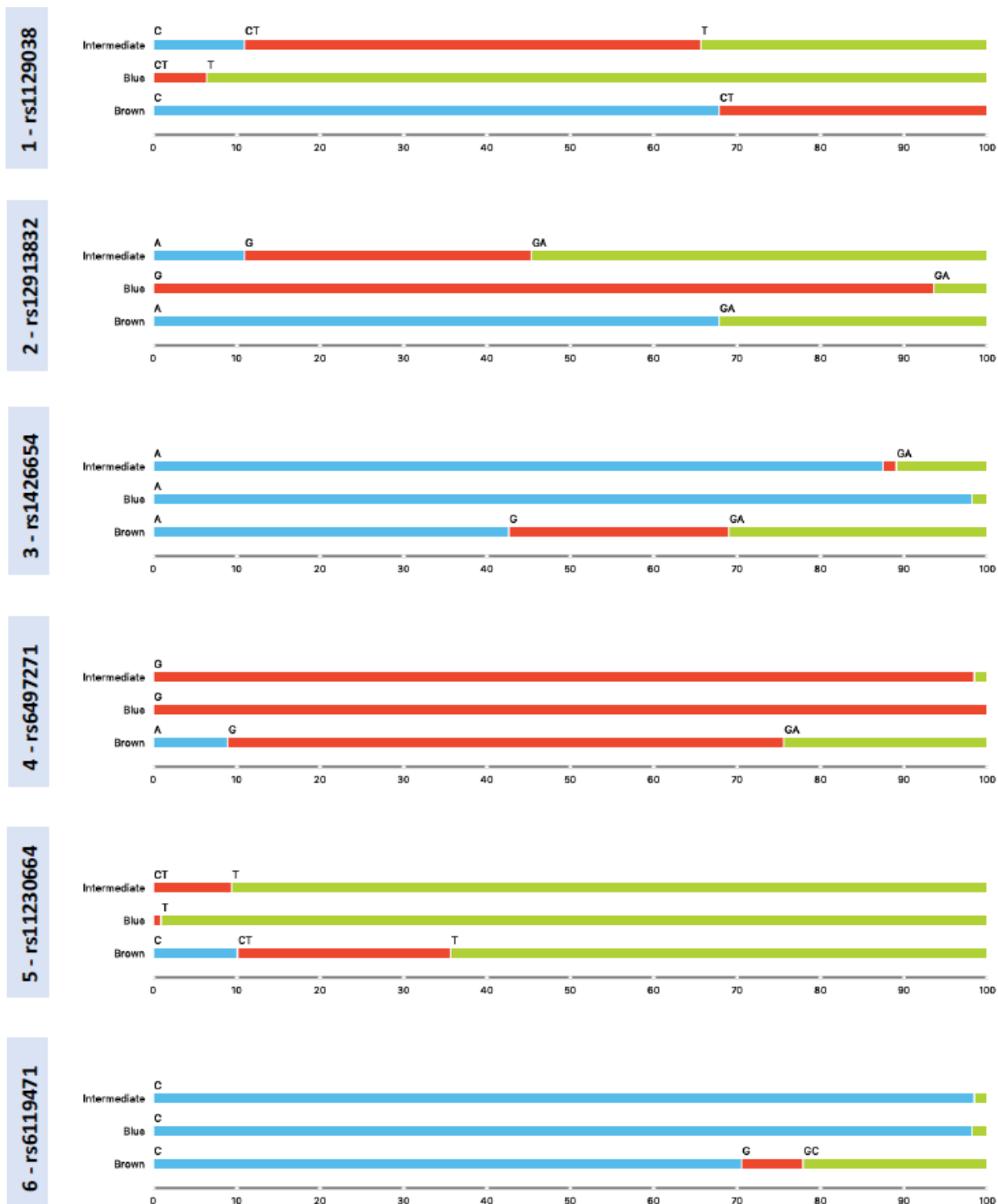


FIGURE S23A – SNPs individual contribution in three eye phenotypic classes (Blue, Intermediate and Brown), in descending importance order.

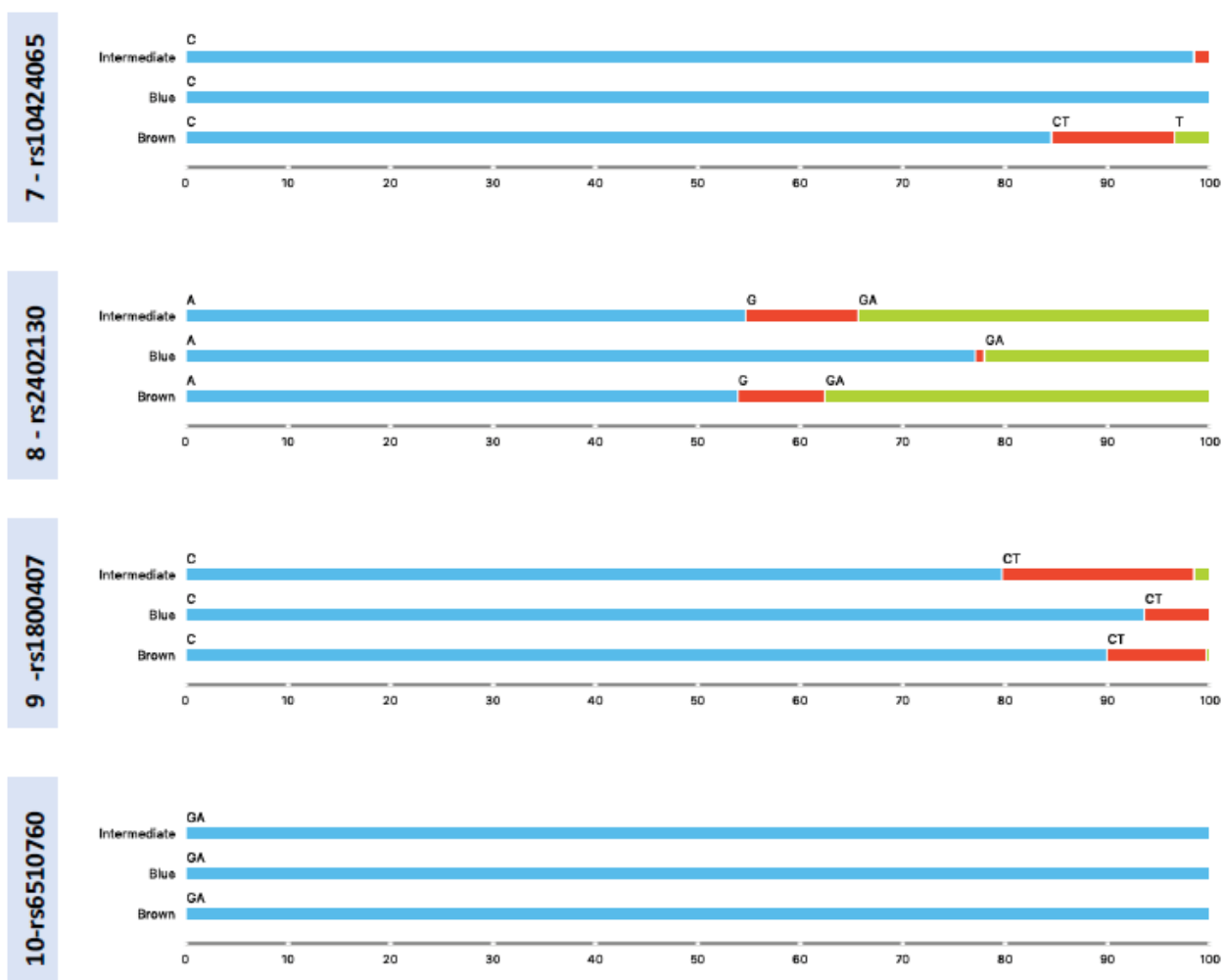


FIGURE S23B – SNPs individual contribution in three eye phenotypic classes (Blue, Intermediate and Brown), in descending importance order.

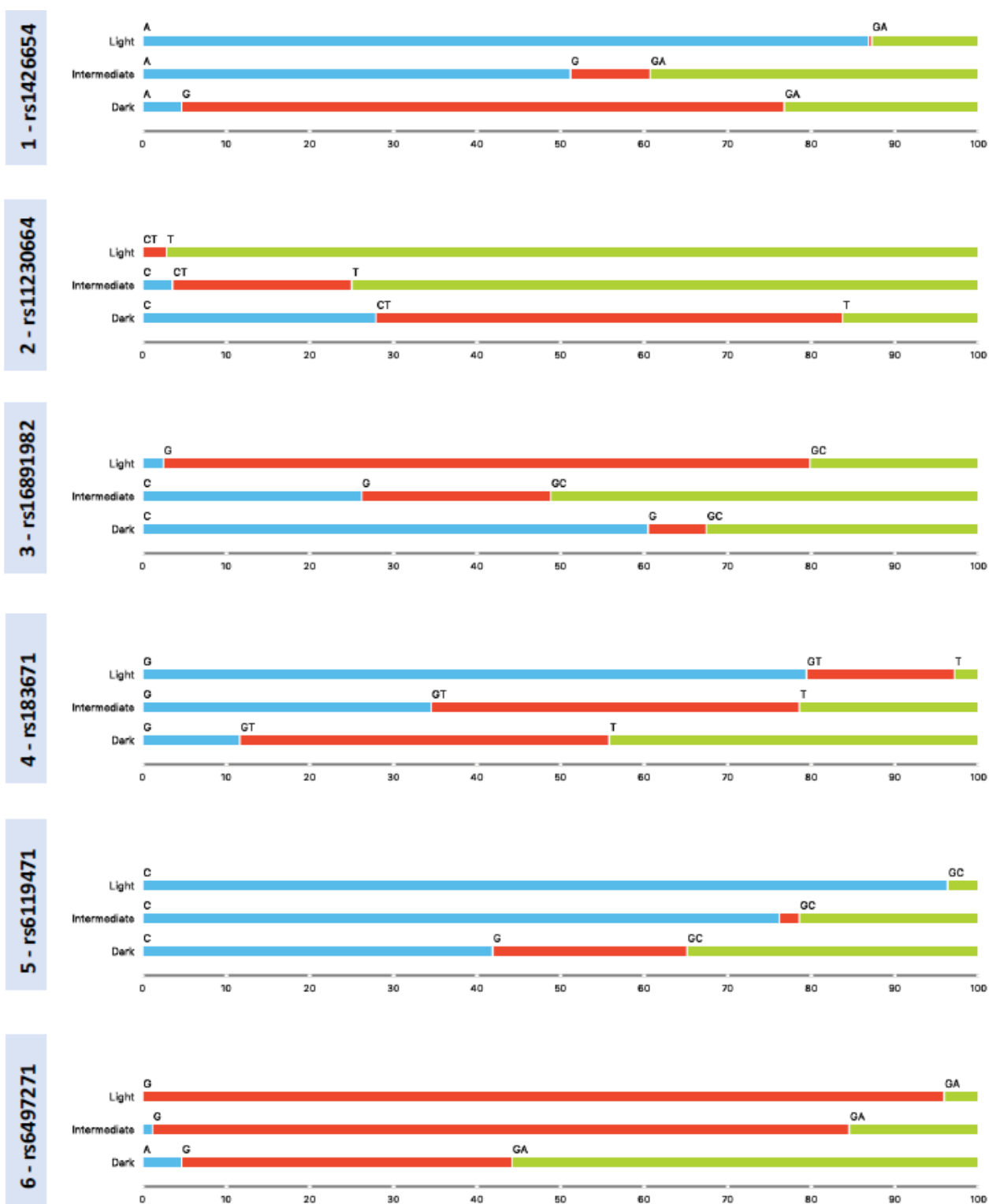


FIGURE S24A – SNPs individual contribution in three skin phenotypic classes (Light, Intermediate and Dark), in descending importance order.

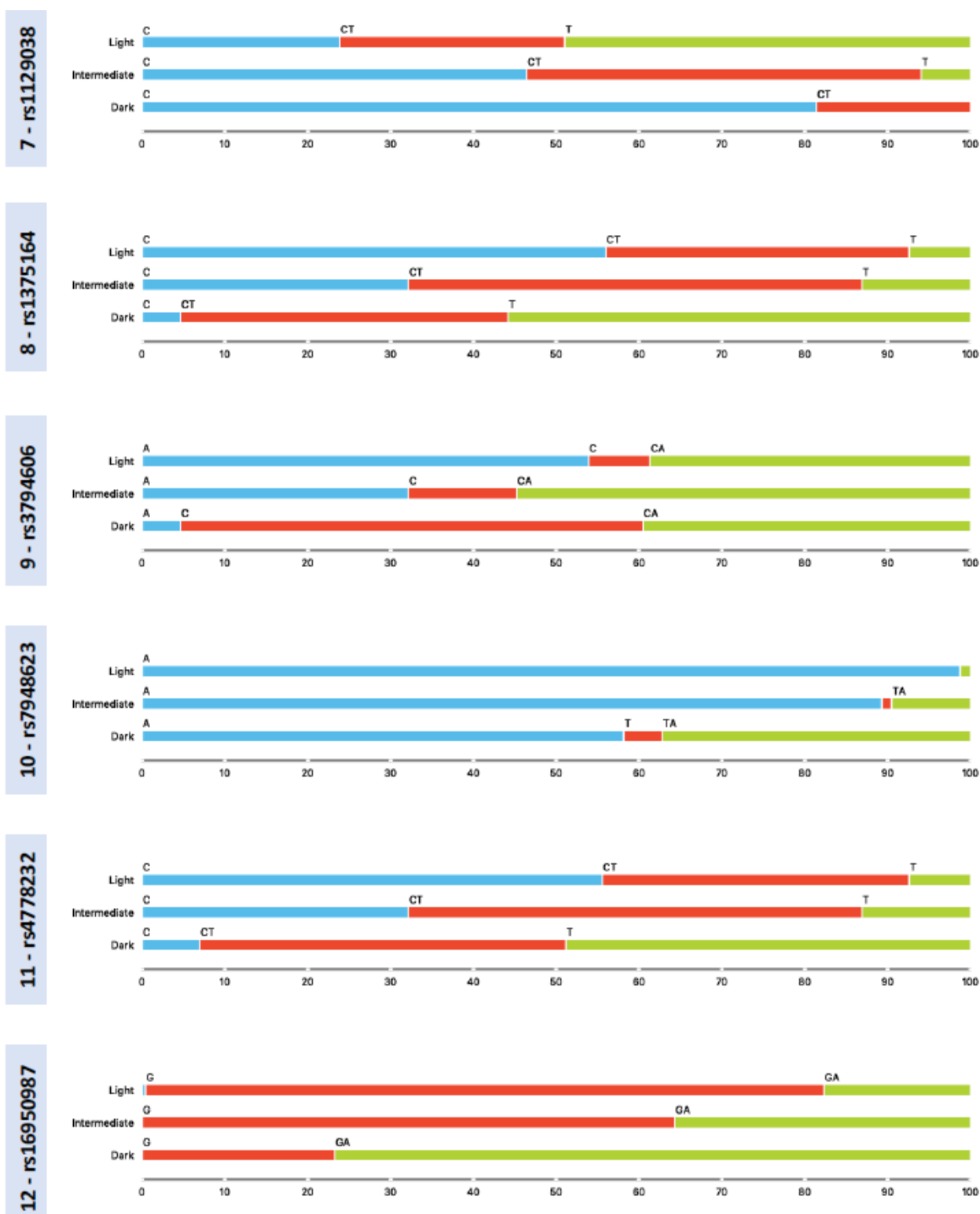


FIGURE S24B – SNPs individual contribution in three skin phenotypic classes (Light, Intermediate and Dark), in descending importance order.

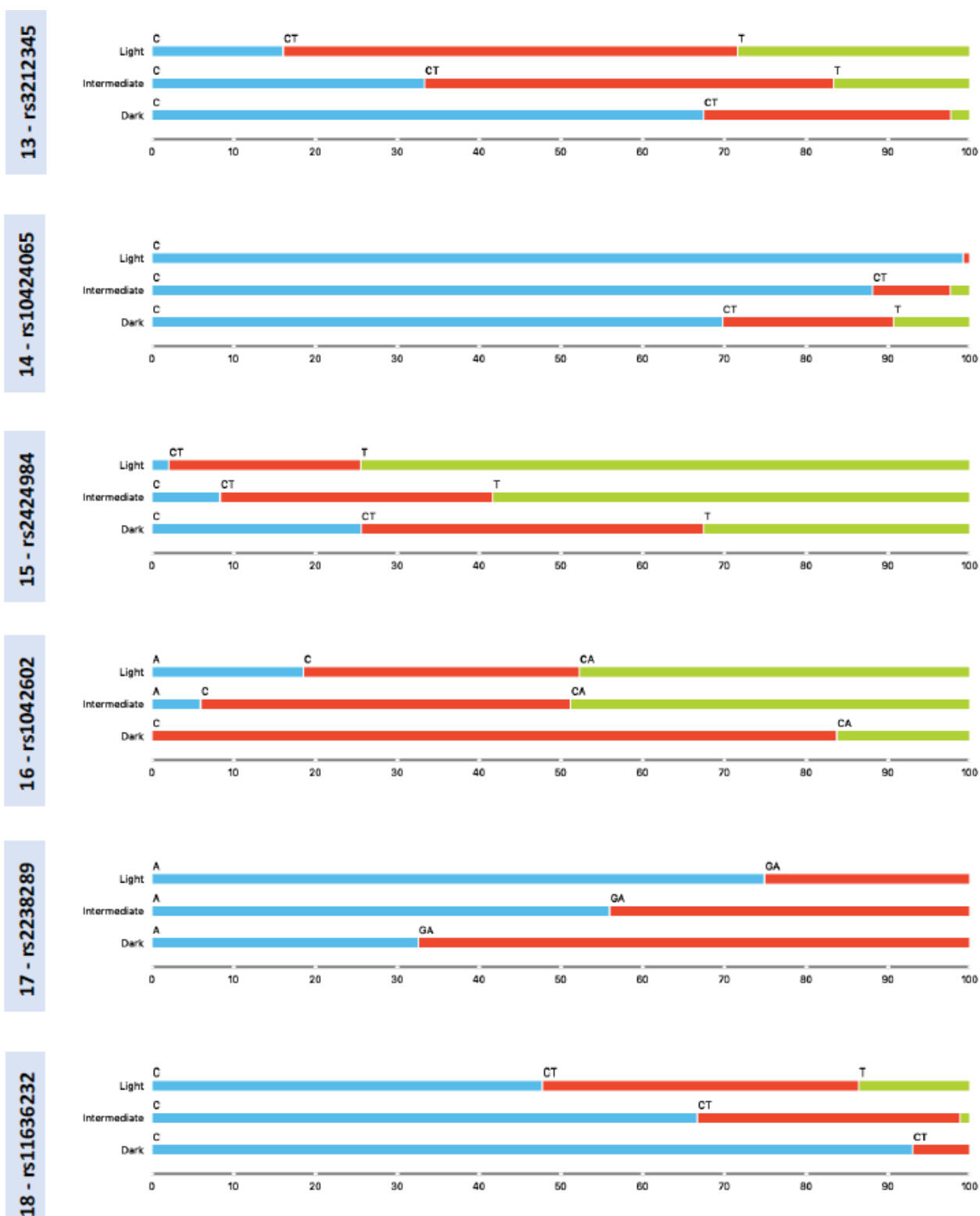


FIGURE S24C – SNPs individual contribution in three skin phenotypic classes (Light, Intermediate and Dark), in descending importance order.

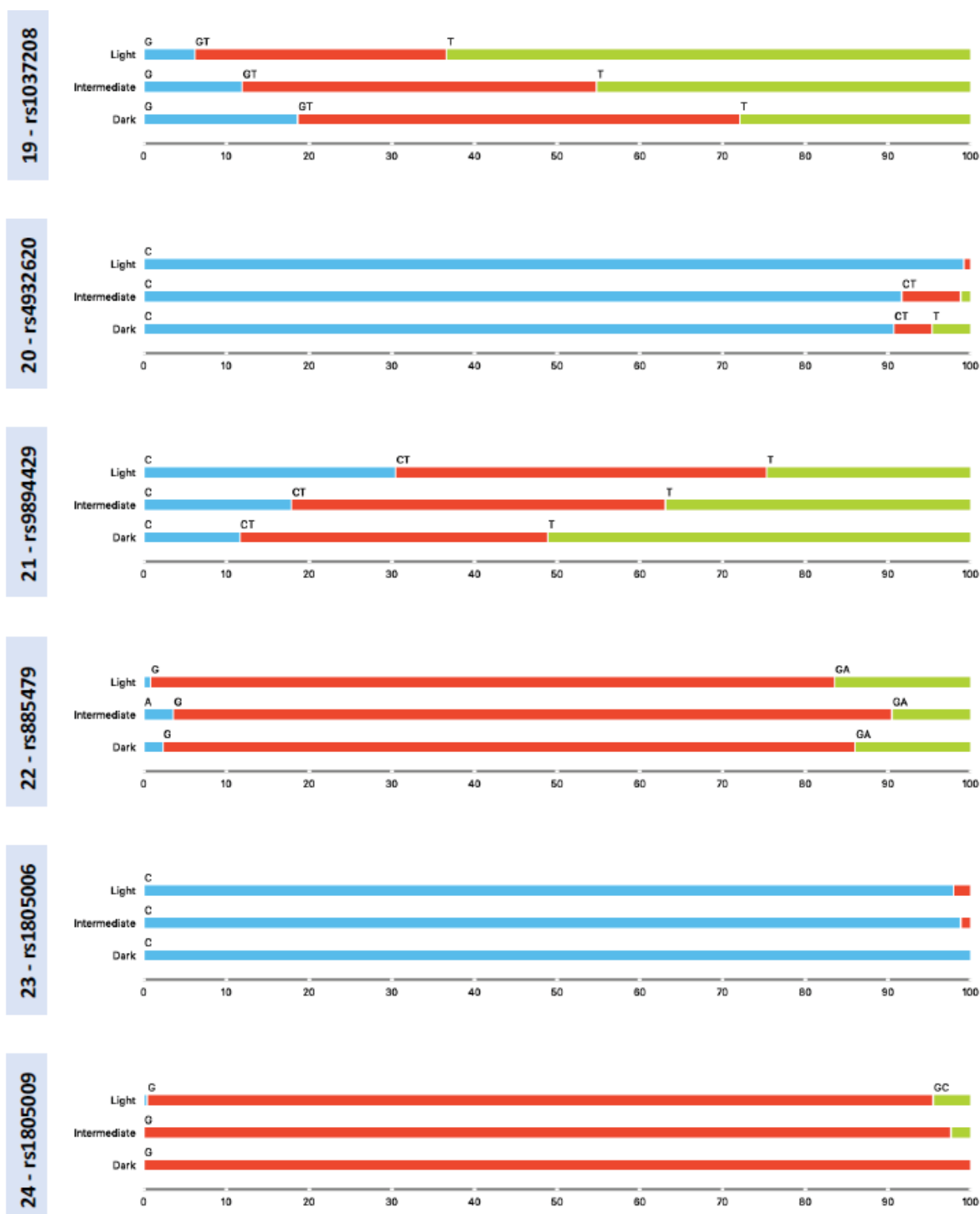


FIGURE S24D – SNPs individual contribution in three skin phenotypic classes (Light, Intermediate and Dark), in descending importance order.

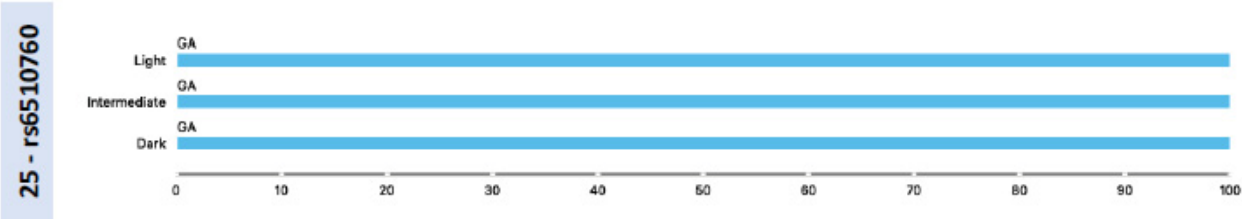


FIGURE S24E – SNPs individual contribution in three skin phenotypic classes (Light, Intermediate and Dark), in descending importance order.

TABLE S1: RFE SNPS SELECTION – EYE – Machine learning models (SVM, RF and MLP) average accuracy (avg) and standard deviation (std) of ten-fold cross validation.

SNP	EYE-31						EYE-32						EYE-33						EYE-4						EYE-5					
	Blue, Green+Hazel,LBrown+DBrown						Blue, Green+Hazel+LBrown, DBrown						Blue+Green, Hazel+LBrown, DBrown						Blue, Green, Hazel, LBrown+DBrown						Blue, Green, Hazel, LBrown, DBrown					
	SVM		RF		MLP		SVM		RF		MLP		SVM		RF		MLP		SVM		RF		MLP		SVM		RF		MLP	
	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std
2	0,83	0,02	0,83	0,02	0,83	0,02	0,79	0,05	0,79	0,05	0,79	0,05	0,79	0,05	0,81	0,06	0,81	0,06	0,83	0,01	0,83	0,01	0,84	0,02	0,74	0,01	0,74	0,02	0,74	0,02
3	0,83	0,02	0,84	0,03	0,83	0,02	0,79	0,06	0,8	0,06	0,8	0,06	0,79	0,07	0,79	0,07	0,81	0,07	0,83	0,02	0,83	0,02	0,83	0,02	0,74	0,02	0,74	0,02	0,74	0,02
4	0,83	0,01	0,84	0,02	0,84	0,02	0,78	0,06	0,79	0,06	0,79	0,06	0,79	0,05	0,8	0,05	0,8	0,05	0,83	0,02	0,83	0,02	0,83	0,01	0,74	0,03	0,74	0,02	0,73	0,02
5	0,84	0,02	0,84	0,02	0,83	0,02	0,78	0,06	0,79	0,06	0,8	0,06	0,79	0,05	0,8	0,07	0,79	0,06	0,83	0,03	0,83	0,03	0,84	0,02	0,74	0,01	0,74	0,01	0,73	0,02
6	0,83	0,01	0,84	0,02	0,84	0,02	0,79	0,05	0,79	0,05	0,79	0,05	0,8	0,06	0,81	0,06	0,81	0,06	0,83	0,02	0,83	0,03	0,83	0,03	0,73	0,01	0,74	0,01	0,74	0,01
7	0,84	0,03	0,85	0,04	0,85	0,03	0,78	0,04	0,76	0,04	0,78	0,04	0,79	0,05	0,8	0,04	0,8	0,05	0,83	0,02	0,83	0,03	0,83	0,02	0,73	0,02	0,74	0,02	0,73	0,03
8	0,85	0,02	0,85	0,03	0,85	0,02	0,81	0,04	0,81	0,04	0,78	0,05	0,81	0,05	0,8	0,05	0,78	0,05	0,83	0,03	0,83	0,03	0,83	0,03	0,74	0,02	0,74	0,02	0,73	0,03
9	0,84	0,02	0,85	0,04	0,85	0,04	0,81	0,04	0,81	0,04	0,79	0,03	0,84	0,08	0,82	0,07	0,82	0,07	0,84	0,02	0,83	0,03	0,83	0,02	0,74	0,02	0,73	0,02	0,73	0,04
10	0,84	0,02	0,86	0,04	0,85	0,03	0,81	0,06	0,79	0,06	0,77	0,06	0,83	0,06	0,81	0,05	0,8	0,07	0,83	0,01	0,84	0,02	0,85	0,03	0,74	0,02	0,72	0,05	0,72	0,05
11	0,84	0,03	0,83	0,04	0,83	0,05	0,81	0,06	0,78	0,04	0,78	0,04	0,83	0,05	0,82	0,05	0,8	0,05	0,83	0,01	0,83	0,01	0,84	0,02	0,74	0,01	0,74	0,01	0,72	0,04
12	0,83	0,03	0,85	0,03	0,84	0,03	0,81	0,05	0,79	0,04	0,79	0,05	0,82	0,05	0,82	0,05	0,83	0,05	0,83	0,03	0,85	0,04	0,82	0,03	0,74	0,02	0,74	0,02	0,72	0,06
13	0,83	0,03	0,83	0,02	0,83	0,03	0,79	0,05	0,79	0,03	0,78	0,04	0,83	0,05	0,81	0,04	0,8	0,05	0,83	0,01	0,83	0,03	0,81	0,03	0,74	0,02	0,74	0,02	0,72	0,04
14	0,83	0,02	0,84	0,04	0,83	0,05	0,8	0,06	0,77	0,05	0,77	0,07	0,82	0,05	0,81	0,05	0,8	0,08	0,83	0,03	0,82	0,03	0,82	0,03	0,74	0,02	0,74	0,02	0,73	0,04
15	0,83	0,03	0,83	0,03	0,83	0,03	0,8	0,06	0,79	0,06	0,78	0,05	0,83	0,05	0,81	0,05	0,79	0,04	0,83	0,02	0,82	0,03	0,82	0,05	0,74	0,03	0,74	0,03	0,73	0,05
16	0,83	0,03	0,83	0,04	0,82	0,06	0,8	0,04	0,8	0,05	0,77	0,07	0,82	0,05	0,82	0,06	0,8	0,04	0,83	0,02	0,82	0,03	0,82	0,04	0,74	0,02	0,73	0,03	0,74	0,04
17	0,83	0,03	0,83	0,04	0,81	0,05	0,79	0,06	0,78	0,04	0,77	0,07	0,82	0,05	0,83	0,06	0,82	0,04	0,83	0,02	0,83	0,02	0,82	0,02	0,74	0,01	0,73	0,03	0,73	0,03
18	0,83	0,03	0,84	0,03	0,8	0,05	0,78	0,03	0,8	0,03	0,76	0,04	0,81	0,06	0,82	0,05	0,81	0,06	0,83	0,02	0,83	0,03	0,82	0,04	0,74	0,03	0,74	0,03	0,71	0,04
19	0,82	0,02	0,83	0,03	0,79	0,05	0,79	0,07	0,79	0,06	0,75	0,07	0,82	0,05	0,82	0,04	0,79	0,07	0,83	0,03	0,83	0,03	0,81	0,06	0,74	0,02	0,74	0,03	0,72	0,02
20	0,83	0,02	0,82	0,03	0,83	0,02	0,8	0,05	0,8	0,03	0,75	0,08	0,83	0,06	0,81	0,03	0,81	0,03	0,83	0,03	0,82	0,04	0,82	0,03	0,74	0,02	0,74	0,03	0,7	0,04
21	0,83	0,02	0,82	0,02	0,8	0,05	0,81	0,06	0,81	0,06	0,78	0,05	0,82	0,06	0,82	0,06	0,81	0,06	0,84	0,02	0,82	0,03	0,81	0,04	0,74	0,03	0,73	0,03	0,69	0,04
22	0,82	0,02	0,82	0,04	0,8	0,07	0,8	0,04	0,78	0,05	0,74	0,04	0,83	0,06	0,81	0,04	0,79	0,06	0,84	0,02	0,83	0,03	0,81	0,04	0,74	0,03	0,75	0,05	0,68	0,08
23	0,83	0,03	0,82	0,03	0,81	0,03	0,79	0,05	0,78	0,07	0,75	0,05	0,81	0,04	0,84	0,05	0,82	0,04	0,84	0,02	0,84	0,02	0,82	0,04	0,75	0,03	0,74	0,04	0,7	0,05
24	0,83	0,03	0,83	0,04	0,78	0,05	0,79	0,04	0,79	0,05	0,76	0,06	0,83	0,05	0,84	0,03	0,78	0,04	0,86	0,04	0,85	0,04	0,85	0,05	0,75	0,03	0,74	0,03	0,69	0,03
25	0,82	0,03	0,82	0,03	0,8	0,05	0,79	0,05	0,77	0,04	0,75	0,04	0,8	0,04	0,82	0,03	0,81	0,04	0,86	0,04	0,85	0,04	0,83	0,04	0,75	0,03	0,73	0,03	0,74	0,04
26	0,84	0,03	0,84	0,03	0,79	0,05	0,78	0,05	0,79	0,05	0,72	0,06	0,82	0,07	0,82	0,08	0,79	0,07	0,86	0,03	0,85	0,03	0,82	0,04	0,75	0,03	0,74	0,04	0,72	0,04
27	0,84	0,03	0,84	0,03	0,78	0,05	0,79	0,05	0,79	0,04	0,76	0,06	0,81	0,04	0,83	0,04	0,79	0,03	0,86	0,03	0,85	0,03	0,81	0,03	0,76	0,04	0,76	0,04	0,71	0,04
28	0,81	0,04	0,83	0,05	0,78	0,07	0,77	0,05	0,79	0,04	0,72	0,06	0,82	0,06	0,84	0,06	0,81	0,04	0,86	0,02	0,85	0,03	0,82	0,04	0,76	0,03	0,74	0,03	0,71	0,07
29	0,84	0,04	0,83	0,03	0,8	0,07	0,78	0,06	0,78	0,06	0,75	0,06	0,8	0,03	0,83	0,05	0,8	0,06	0,86	0,03	0,86	0,03	0,83	0,05	0,75	0,03	0,75	0,03	0,71	0,07
30	0,84	0,03	0,85	0,02	0,82	0,04	0,77	0,04	0,79	0,05	0,74	0,05	0,8	0,04	0,83	0,05	0,82	0,07	0,86	0,04	0,86	0,04	0,8	0,05	0,75	0,03	0,73	0,03	0,72	0,06
31	0,84	0,03	0,83	0,05	0,77	0,07	0,79	0,05	0,79	0,05	0,75	0,05	0,81	0,04	0,81	0,06	0,81	0,04	0,86	0,03	0,85	0,04	0,83	0,04	0,75	0,02	0,74	0,02	0,71	0,04
32	0,82	0,04	0,83	0,03	0,8	0,04	0,77	0,05	0,79	0,04	0,72	0,06	0,81	0,04	0,82	0,04	0,78	0,06	0,86	0,02	0,85	0,02	0,77	0,06	0,75	0,03	0,74	0,03	0,69	0,06
33	0,83	0,04	0,83	0,04	0,79	0,06	0,79	0,06	0,79	0,07	0,74	0,03	0,82	0,04	0,83	0,04	0,78	0,06	0,86	0,03	0,85	0,03	0,81	0,04	0,76	0,03	0,75	0,05	0,67	0,09
34	0,82	0,04	0,83	0,04	0,77	0,06	0,79	0,06	0,79	0,07	0,76	0,05	0,79	0,04	0,82	0,05	0,78	0,04	0,86	0,04	0,85	0,04	0,78	0,05	0,76	0,05	0,76	0,05	0,71	0,05
35	0,82	0,02	0,82	0,03	0,78	0,06	0,79	0,04	0,79	0,04	0,74	0,05	0,79	0,04	0,8	0,04	0,75	0,1	0,86	0,04	0,85	0,04	0,8	0,05	0,75	0,02	0,76	0,02	0,7	0,05
36	0,85	0,03	0,83	0,03	0,79	0,06	0,79	0,06	0,79	0,06	0,72	0,08	0,79	0,04	0,82	0,05	0,79	0,06	0,86	0,04	0,85	0,03	0,8	0,05	0,78	0,04	0,78	0,03	0,73	0,05
37	0,85	0,05	0,85	0,05	0,81	0,07	0,79	0,04	0,81	0,03	0,75	0,04	0,8	0,04	0,81	0,04	0,77	0,06	0,86	0,03	0,85	0,03	0,83	0,04	0,77	0,06	0,77	0,02	0,68	0,07
38	0,84	0,04	0,84	0,04	0,8	0,06	0,76	0,07	0,8	0,05	0,74	0,07	0,82	0,04	0,81	0,05	0,7													

TABLE S2: X^2 SNPS SELECTION – EYE – Machine learning models (SVM, RF and MLP) average accuracy (avg) and standard deviation (std) of ten-fold cross validation.

SNP	EYE-31						EYE-32						EYE-33						EYE-4						EYE-5					
	Blue, Green+Hazel, LBrown+DBrown						Blue, Green+Hazel+LBrown, DBrown						Blue+Green, Hazel+LBrown, DBrown						Blue, Green, Hazel, LBrown+DBrown						Blue, Green, Hazel, LBrown, DBrown					
	SVM		RF		MLP		SVM		RF		MLP		SVM		RF		MLP		SVM		RF		MLP		SVM		RF		MLP	
	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std
2	0.83	0.02	0.83	0.02	0.83	0.02	0.74	0.05	0.76	0.06	0.76	0.06	0.79	0.04	0.79	0.03	0.79	0.03	0.83	0.01	0.83	0.01	0.83	0.01	0.74	0.02	0.74	0.02	0.74	0.02
3	0.83	0.02	0.84	0.02	0.83	0.02	0.82	0.04	0.82	0.04	0.82	0.04	0.84	0.04	0.84	0.04	0.84	0.04	0.83	0.02	0.83	0.02	0.83	0.02	0.74	0.02	0.73	0.02	0.73	0.03
4	0.83	0.02	0.84	0.02	0.83	0.03	0.81	0.07	0.8	0.05	0.82	0.06	0.83	0.05	0.82	0.05	0.83	0.05	0.83	0.02	0.83	0.02	0.83	0.03	0.75	0.02	0.75	0.02	0.76	0.04
5	0.83	0.02	0.84	0.03	0.83	0.04	0.82	0.05	0.81	0.04	0.8	0.03	0.82	0.04	0.82	0.05	0.83	0.04	0.83	0.01	0.83	0.02	0.82	0.02	0.75	0.03	0.75	0.03	0.75	0.03
6	0.84	0.02	0.82	0.02	0.83	0.03	0.81	0.06	0.8	0.04	0.8	0.06	0.81	0.04	0.81	0.04	0.81	0.04	0.84	0.02	0.83	0.02	0.83	0.02	0.75	0.03	0.75	0.04	0.75	0.03
7	0.84	0.02	0.82	0.03	0.82	0.02	0.8	0.04	0.8	0.03	0.8	0.04	0.81	0.05	0.81	0.04	0.82	0.06	0.84	0.03	0.84	0.03	0.83	0.03	0.76	0.04	0.76	0.04	0.75	0.03
8	0.84	0.02	0.83	0.03	0.83	0.04	0.81	0.05	0.79	0.04	0.79	0.04	0.81	0.04	0.81	0.03	0.84	0.06	0.84	0.02	0.83	0.03	0.83	0.03	0.77	0.05	0.77	0.04	0.76	0.04
9	0.83	0.03	0.83	0.03	0.83	0.04	0.8	0.04	0.78	0.03	0.78	0.04	0.83	0.04	0.83	0.03	0.82	0.04	0.84	0.03	0.84	0.03	0.82	0.03	0.75	0.03	0.77	0.03	0.76	0.04
10	0.84	0.02	0.83	0.04	0.83	0.04	0.81	0.06	0.79	0.05	0.81	0.07	0.83	0.04	0.83	0.06	0.82	0.04	0.84	0.03	0.84	0.03	0.81	0.04	0.74	0.03	0.77	0.04	0.73	0.04
11	0.84	0.03	0.84	0.03	0.83	0.04	0.8	0.07	0.77	0.05	0.79	0.06	0.85	0.06	0.85	0.06	0.83	0.05	0.84	0.03	0.82	0.05	0.82	0.05	0.75	0.03	0.78	0.04	0.76	0.04
12	0.84	0.03	0.85	0.04	0.83	0.05	0.8	0.07	0.8	0.06	0.8	0.06	0.82	0.04	0.83	0.04	0.78	0.11	0.84	0.02	0.83	0.03	0.83	0.03	0.76	0.03	0.78	0.04	0.75	0.05
13	0.84	0.03	0.83	0.04	0.83	0.05	0.78	0.06	0.8	0.05	0.8	0.03	0.83	0.04	0.82	0.05	0.81	0.05	0.84	0.02	0.84	0.03	0.81	0.05	0.76	0.03	0.78	0.03	0.75	0.04
14	0.84	0.02	0.84	0.05	0.83	0.05	0.79	0.06	0.81	0.05	0.81	0.06	0.83	0.04	0.83	0.04	0.82	0.05	0.84	0.03	0.83	0.03	0.83	0.03	0.76	0.02	0.77	0.05	0.74	0.05
15	0.84	0.02	0.84	0.03	0.83	0.03	0.8	0.03	0.8	0.04	0.79	0.07	0.84	0.04	0.83	0.04	0.82	0.05	0.84	0.02	0.83	0.02	0.82	0.03	0.75	0.04	0.77	0.04	0.76	0.04
16	0.84	0.02	0.83	0.01	0.82	0.06	0.81	0.06	0.8	0.05	0.81	0.05	0.84	0.05	0.82	0.06	0.82	0.03	0.84	0.02	0.84	0.02	0.8	0.04	0.76	0.03	0.77	0.04	0.76	0.06
17	0.84	0.03	0.83	0.05	0.8	0.05	0.8	0.05	0.79	0.03	0.8	0.05	0.83	0.04	0.83	0.05	0.8	0.06	0.84	0.03	0.84	0.03	0.81	0.06	0.76	0.04	0.77	0.04	0.74	0.05
18	0.84	0.02	0.84	0.03	0.8	0.06	0.81	0.04	0.8	0.04	0.79	0.05	0.84	0.04	0.83	0.04	0.81	0.05	0.84	0.02	0.84	0.02	0.81	0.03	0.76	0.02	0.76	0.05	0.76	0.05
19	0.84	0.02	0.84	0.03	0.8	0.05	0.81	0.04	0.8	0.04	0.78	0.07	0.84	0.03	0.82	0.04	0.8	0.04	0.84	0.02	0.83	0.03	0.81	0.04	0.76	0.02	0.76	0.03	0.75	0.04
20	0.84	0.02	0.83	0.04	0.81	0.05	0.79	0.09	0.78	0.08	0.77	0.09	0.83	0.05	0.83	0.05	0.8	0.08	0.84	0.02	0.83	0.02	0.78	0.04	0.76	0.02	0.77	0.04	0.74	0.03
21	0.84	0.01	0.83	0.03	0.79	0.05	0.81	0.04	0.79	0.05	0.78	0.08	0.83	0.04	0.84	0.03	0.8	0.04	0.84	0.02	0.83	0.02	0.8	0.05	0.76	0.03	0.77	0.04	0.75	0.04
22	0.84	0.02	0.84	0.03	0.81	0.04	0.79	0.06	0.79	0.04	0.74	0.07	0.84	0.04	0.83	0.04	0.78	0.04	0.84	0.03	0.83	0.03	0.78	0.05	0.76	0.03	0.75	0.04	0.7	0.07
23	0.84	0.02	0.83	0.02	0.79	0.04	0.79	0.04	0.8	0.03	0.75	0.06	0.84	0.05	0.84	0.05	0.8	0.03	0.84	0.03	0.84	0.03	0.79	0.05	0.76	0.02	0.76	0.03	0.72	0.04
24	0.84	0.03	0.84	0.03	0.79	0.07	0.81	0.03	0.79	0.01	0.79	0.04	0.84	0.04	0.84	0.05	0.82	0.03	0.84	0.02	0.83	0.02	0.79	0.04	0.76	0.02	0.76	0.04	0.7	0.05
25	0.84	0.02	0.82	0.03	0.77	0.06	0.78	0.05	0.8	0.04	0.72	0.05	0.82	0.05	0.82	0.07	0.76	0.04	0.83	0.03	0.83	0.03	0.79	0.05	0.76	0.02	0.75	0.04	0.74	0.04
26	0.85	0.02	0.84	0.02	0.81	0.04	0.78	0.06	0.76	0.06	0.75	0.05	0.82	0.04	0.83	0.04	0.77	0.03	0.83	0.03	0.83	0.04	0.76	0.06	0.76	0.02	0.76	0.04	0.73	0.06
27	0.85	0.02	0.84	0.03	0.78	0.05	0.77	0.06	0.79	0.06	0.73	0.1	0.82	0.04	0.83	0.04	0.74	0.05	0.84	0.02	0.83	0.02	0.78	0.06	0.76	0.02	0.76	0.03	0.74	0.04
28	0.84	0.03	0.84	0.05	0.79	0.06	0.79	0.07	0.77	0.06	0.74	0.06	0.82	0.02	0.82	0.03	0.75	0.06	0.84	0.02	0.83	0.03	0.81	0.04	0.73	0.03	0.74	0.05	0.66	0.06
29	0.84	0.03	0.83	0.04	0.8	0.03	0.78	0.04	0.78	0.06	0.78	0.05	0.81	0.06	0.83	0.07	0.8	0.06	0.84	0.02	0.84	0.03	0.79	0.04	0.74	0.03	0.74	0.03	0.71	0.05
30	0.84	0.03	0.81	0.04	0.79	0.06	0.78	0.07	0.78	0.07	0.72	0.06	0.81	0.03	0.82	0.04	0.8	0.08	0.84	0.04	0.83	0.06	0.81	0.05	0.74	0.03	0.73	0.03	0.7	0.05
31	0.83	0.02	0.83	0.03	0.77	0.05	0.77	0.08	0.8	0.09	0.75	0.07	0.81	0.06	0.81	0.05	0.79	0.04	0.82	0.03	0.82	0.03	0.77	0.05	0.74	0.02	0.75	0.04	0.73	0.04
32	0.84	0.03	0.83	0.04	0.78	0.05	0.78	0.06	0.79	0.07	0.77	0.05	0.81	0.05	0.82	0.06	0.8	0.07	0.82	0.03	0.82	0.03	0.75	0.06	0.74	0.03	0.73	0.03	0.73	0.06
33	0.84	0.03	0.83	0.03	0.78	0.03	0.76	0.03	0.79	0.04	0.73	0.05	0.81	0.05	0.82	0.06	0.8	0.06	0.83	0.03	0.81	0.03	0.78	0.06	0.74	0.02	0.74	0.05	0.71	0.07
34	0.84	0.03	0.84	0.03	0.76	0.09	0.78	0.05	0.78	0.05	0.72	0.08	0.82	0.07	0.8	0.06	0.78	0.08	0.82	0.02	0.81	0.03	0.76	0.07	0.75	0.02	0.73	0.03	0.67	0.06
35	0.84	0.03	0.83	0.03	0.74	0.04	0.77	0.06	0.78	0.07	0.7	0.09	0.8	0.05	0.82	0.06	0.79	0.06	0.82	0.02	0.83	0.03	0.77	0.08	0.74	0.02	0.75	0.04	0.71	0.05
36	0.83	0.03	0.83	0.03	0.79	0.05	0.77	0.03	0.77	0.04	0.7	0.06	0.8	0.05	0.8	0.05	0.8	0.06	0.82	0.03	0.81	0.04	0.75	0.05	0.74	0.03	0.73	0.07	0.68	0.05
37	0.84	0.03	0.82	0.03	0.76	0.07	0.78	0.05	0.78	0.06	0.68	0.06	0.8	0.06	0.8	0.07	0.78	0.05	0.83	0.03	0.81	0.04	0.72	0.05	0.74	0.02	0.75	0.05	0.71	0.08
38	0.85	0.02	0.83	0.03	0.78	0.06	0.77	0.07	0.76	0.05	0.7	0.08	0.8	0.06	0.81	0.04	0.79	0.06	0.82	0.02	0.82	0.04	0.78	0.06	0.73	0.03	0.74	0.03	0.71	0.05
39	0.84	0.04	0.83	0.05	0.79	0.05	0.77	0.08	0.78	0.07	0.68	0.1	0.81	0.06	0.8	0.06	0.74	0.12	0.83	0.02	0.81	0.05	0.77	0.04	0.74	0.03	0.75	0.04	0.68	0.06
40	0.84	0.02	0.82	0.05	0.77	0.03	0.77	0.05	0.77	0.04	0.7	0.05	0.8	0.06	0.81	0.04	0.78	0.04	0.83	0.02	0.82	0.02	0.74	0.05	0.74	0.03	0.73	0.04	0.67	0.06
41	0.84	0.03	0.83	0.04	0.79	0.05	0.78	0.05	0.77	0.06	0.71	0.07	0.8	0.05	0.8	0.04	0.8	0.05	0.83	0.03	0.81	0.03	0.76	0.06	0.74	0.03	0.74	0.04	0.66	0.06
42	0.83	0.02	0.83	0.03	0.77	0.06	0.77	0.06	0.78	0.06	0.68	0.07	0.82	0.03	0.8	0.03	0.75	0.12	0.83	0.02	0.81	0.03	0.75	0.08	0.74	0.02	0.75	0.03	0.67	0.06
43	0.83	0.02	0.83	0.02	0.75	0.06	0.79	0.05	0.77	0.07	0.7	0.07	0.79	0.04	0.79	0.07	0.79	0.04	0.82	0.04	0.82	0.04	0.75	0.05	0.74	0.04	0.73	0.04	0.74	0.06
44	0.86	0.02	0.84	0.03	0.78	0.05	0.77	0.09	0.77	0.08	0.7	0.09	0.79	0.06	0.79															

TABLE S3: RFE SNPS SELECTION – SKIN 234 – Machine learning models (SVM, RF and MLP) average accuracy (avg) and standard deviation (std) of ten-fold cross validation.

SNP	SKYN – 2						SKYN – 3						SKYN – 4					
	White+Pale+Beige, LBrown+MBrown+DBrown						White+Pale, Beige+LBrown, MBrown+DBrown						White+Pale, Beige, LBrown, MBrown+DBrown					
	SVM		RF		MLP		SVM		RF		MLP		SVM		RF		MLP	
	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std
2	0.89	0.03	0.9	0.03	0.9	0.03	0.72	0.03	0.73	0.04	0.72	0.03	0.72	0.03	0.72	0.03	0.72	0.03
3	0.89	0.04	0.89	0.03	0.89	0.04	0.72	0.03	0.72	0.03	0.72	0.03	0.72	0.02	0.72	0.03	0.73	0.03
4	0.89	0.04	0.9	0.03	0.91	0.04	0.72	0.03	0.74	0.05	0.74	0.03	0.72	0.04	0.72	0.03	0.72	0.03
5	0.92	0.05	0.92	0.05	0.92	0.04	0.73	0.03	0.73	0.03	0.75	0.04	0.72	0.03	0.72	0.03	0.74	0.02
6	0.92	0.03	0.93	0.02	0.92	0.02	0.72	0.04	0.72	0.04	0.75	0.04	0.73	0.03	0.73	0.03	0.74	0.03
7	0.92	0.03	0.92	0.02	0.93	0.02	0.73	0.04	0.73	0.02	0.73	0.03	0.72	0.03	0.73	0.02	0.72	0.03
8	0.92	0.03	0.93	0.03	0.92	0.02	0.73	0.07	0.73	0.03	0.73	0.06	0.71	0.05	0.72	0.03	0.72	0.04
9	0.95	0.03	0.94	0.03	0.94	0.03	0.73	0.04	0.73	0.02	0.75	0.06	0.74	0.03	0.75	0.02	0.72	0.05
10	0.95	0.02	0.94	0.02	0.94	0.04	0.72	0.03	0.74	0.02	0.73	0.04	0.75	0.03	0.73	0.02	0.73	0.04
11	0.94	0.03	0.94	0.04	0.94	0.04	0.73	0.04	0.73	0.03	0.72	0.07	0.77	0.03	0.76	0.04	0.77	0.05
12	0.94	0.03	0.94	0.04	0.94	0.03	0.75	0.02	0.75	0.01	0.74	0.04	0.77	0.04	0.75	0.02	0.76	0.04
13	0.95	0.02	0.95	0.04	0.93	0.03	0.75	0.03	0.74	0.03	0.76	0.07	0.77	0.03	0.75	0.03	0.77	0.08
14	0.95	0.02	0.96	0.03	0.93	0.04	0.75	0.02	0.76	0.02	0.77	0.05	0.78	0.02	0.75	0.03	0.76	0.04
15	0.95	0.04	0.95	0.05	0.96	0.03	0.75	0.03	0.77	0.02	0.76	0.06	0.77	0.02	0.76	0.04	0.77	0.04
16	0.95	0.04	0.95	0.05	0.93	0.05	0.77	0.05	0.76	0.03	0.75	0.06	0.77	0.03	0.76	0.04	0.75	0.03
17	0.95	0.02	0.96	0.02	0.94	0.04	0.76	0.04	0.74	0.03	0.75	0.07	0.77	0.02	0.76	0.03	0.76	0.03
18	0.96	0.02	0.95	0.02	0.94	0.03	0.78	0.06	0.74	0.02	0.72	0.04	0.77	0.03	0.74	0.04	0.74	0.04
19	0.94	0.04	0.94	0.03	0.92	0.05	0.77	0.05	0.77	0.03	0.72	0.06	0.77	0.04	0.77	0.02	0.73	0.06
20	0.95	0.03	0.94	0.03	0.93	0.03	0.76	0.02	0.75	0.05	0.77	0.06	0.77	0.04	0.75	0.04	0.72	0.05
21	0.95	0.03	0.94	0.03	0.91	0.05	0.75	0.02	0.76	0.02	0.74	0.04	0.78	0.04	0.75	0.04	0.73	0.04
22	0.94	0.05	0.94	0.04	0.93	0.04	0.75	0.07	0.74	0.04	0.75	0.09	0.77	0.04	0.74	0.05	0.73	0.04
23	0.95	0.03	0.95	0.03	0.93	0.04	0.72	0.07	0.75	0.03	0.71	0.05	0.77	0.03	0.76	0.03	0.76	0.05
24	0.95	0.03	0.95	0.04	0.94	0.04	0.79	0.04	0.77	0.04	0.77	0.05	0.77	0.03	0.77	0.02	0.72	0.04
25	0.95	0.04	0.94	0.04	0.93	0.04	0.81	0.06	0.77	0.05	0.76	0.05	0.77	0.04	0.75	0.04	0.74	0.06
26	0.96	0.02	0.95	0.04	0.95	0.03	0.81	0.04	0.8	0.04	0.73	0.11	0.77	0.03	0.76	0.02	0.73	0.03
27	0.95	0.03	0.94	0.03	0.93	0.03	0.81	0.05	0.77	0.04	0.78	0.06	0.77	0.04	0.76	0.04	0.74	0.06
28	0.96	0.03	0.96	0.03	0.93	0.02	0.81	0.03	0.79	0.05	0.76	0.06	0.77	0.03	0.76	0.03	0.71	0.06
29	0.96	0.02	0.94	0.02	0.92	0.05	0.8	0.05	0.78	0.08	0.77	0.06	0.78	0.03	0.76	0.02	0.68	0.06
30	0.95	0.03	0.95	0.04	0.92	0.06	0.79	0.07	0.77	0.07	0.75	0.1	0.76	0.04	0.75	0.03	0.76	0.06
31	0.96	0.02	0.95	0.03	0.94	0.04	0.81	0.03	0.77	0.05	0.74	0.07	0.77	0.03	0.76	0.04	0.69	0.05
32	0.97	0.02	0.95	0.04	0.95	0.02	0.76	0.06	0.76	0.05	0.76	0.07	0.78	0.03	0.77	0.03	0.74	0.07
33	0.97	0.03	0.95	0.03	0.96	0.03	0.79	0.06	0.79	0.06	0.76	0.05	0.78	0.03	0.76	0.04	0.72	0.06
34	0.96	0.03	0.97	0.03	0.95	0.03	0.8	0.04	0.77	0.04	0.76	0.05	0.77	0.05	0.77	0.06	0.72	0.07
35	0.97	0.03	0.95	0.04	0.95	0.04	0.82	0.06	0.79	0.05	0.77	0.07	0.78	0.04	0.76	0.03	0.77	0.04
36	0.98	0.01	0.94	0.04	0.96	0.02	0.8	0.07	0.79	0.05	0.72	0.06	0.77	0.02	0.73	0.03	0.73	0.06
37	0.97	0.03	0.95	0.03	0.96	0.03	0.8	0.04	0.79	0.06	0.79	0.05	0.76	0.05	0.75	0.03	0.73	0.05
38	0.97	0.03	0.96	0.03	0.94	0.03	0.8	0.06	0.78	0.06	0.77	0.05	0.77	0.04	0.76	0.03	0.7	0.05
39	0.98	0.02	0.94	0.04	0.95	0.05	0.81	0.04	0.78	0.05	0.72	0.05	0.76	0.02	0.75	0.03	0.7	0.05
40	0.98	0.02	0.96	0.02	0.95	0.04	0.79	0.05	0.8	0.03	0.73	0.1	0.77	0.05	0.77	0.04	0.68	0.08
41	0.97	0.03	0.95	0.03	0.95	0.03	0.8	0.04	0.78	0.05	0.77	0.05	0.77	0.04	0.75	0.04	0.72	0.09
42	0.96	0.03	0.95	0.03	0.95	0.02	0.81	0.06	0.75	0.04	0.76	0.06	0.77	0.04	0.77	0.03	0.71	0.06
43	0.97	0.02	0.96	0.03	0.94	0.03	0.82	0.05	0.81	0.05	0.74	0.05	0.78	0.05	0.77	0.06	0.72	0.06
44	0.97	0.02	0.95	0.03	0.94	0.04	0.82	0.04	0.78	0.04	0.74	0.08	0.77	0.04	0.76	0.03	0.71	0.07
45	0.97	0.04	0.95	0.03	0.95	0.04	0.8	0.05	0.8	0.05	0.73	0.04	0.76	0.04	0.75	0.04	0.7	0.07
46	0.96	0.04	0.95	0.03	0.94	0.06	0.79	0.05	0.79	0.06	0.77	0.04	0.78	0.03	0.75	0.05	0.7	0.09
47	0.96	0.03	0.95	0.04	0.93	0.04	0.81	0.05	0.79	0.07	0.76	0.04	0.77	0.03	0.77	0.04	0.7	0.05
48	0.96	0.02	0.94	0.04	0.95	0.02	0.8	0.06	0.8	0.06	0.74	0.05	0.77	0.03	0.75	0.02	0.7	0.06
49	0.96	0.03	0.95	0.03	0.94	0.04	0.8	0.04	0.79	0.05	0.76	0.08	0.78	0.03	0.77	0.04	0.7	0.05
50	0.96	0.02	0.95	0.03	0.94	0.03	0.79	0.07	0.8	0.05	0.71	0.04	0.78	0.04	0.78	0.03	0.71	0.05
51	0.96	0.03	0.96	0.03	0.94	0.05	0.77	0.05	0.77	0.05	0.73	0.05	0.78	0.02	0.8	0.04	0.75	0.03
52	0.96	0.03	0.94	0.04	0.94	0.02	0.8	0.05	0.79	0.05	0.75	0.06	0.79	0.04	0.79	0.05	0.71	0.06
53	0.96	0.03	0.95	0.04	0.94	0.02	0.8	0.05	0.81	0.08	0.7	0.05	0.78	0.05	0.77	0.03	0.72	0.08
54	0.97	0.03	0.95	0.02	0.95	0.03	0.79	0.06	0.75	0.06	0.72	0.06	0.8	0.03	0.78	0.03	0.75	0.04
55	0.96	0.03	0.95	0.04	0.95	0.03	0.82	0.05	0.79	0.06	0.74	0.07	0.79	0.04	0.78	0.05	0.69	0.06
56	0.97	0.03	0.94	0.02	0.95	0.02	0.79	0.05	0.78	0.03	0.74	0.04	0.77	0.03	0.77	0.03	0.7	0.06
57	0.97	0.02	0.94	0.03	0.94	0.04	0.79	0.05	0.79	0.04	0.75	0.08	0.78	0.03	0.77	0.02	0.69	0.05
58	0.95	0.02	0.94	0.03	0.96	0.02	0.8	0.05	0.79	0.05	0.73	0.08	0.79	0.03	0.79	0.04	0.7	0.08
59	0.96	0.03	0.93	0.05	0.95	0.04	0.81	0.04	0.78	0.05	0.76	0.05	0.8	0.03	0.76	0.03	0.78	0.04
60	0.96	0.03	0.94	0.05	0.94	0.04	0.81	0.06	0.8	0.05	0.73	0.06	0.78	0.04	0.78	0.03	0.71	0.07
61	0.97	0.02	0.93	0.03	0.95	0.03	0.8	0.06	0.8	0.05	0.76	0.05	0.79	0.04	0.79	0.05	0.73	0.05
62	0.97	0.04	0.94	0.04	0.95	0.03	0.8	0.07	0.8	0.03	0.78	0.08	0.79	0.06	0.78	0.05	0.72	0.06
63	0.96	0.02	0.95	0.05	0.95	0.02	0.83	0.06	0.79	0.06	0.73	0.08	0.8	0.06	0.8	0.06	0.74	0.08
64	0.95	0.04	0.94	0.05	0.94	0.05	0.8	0.03	0.81	0.05	0.74	0.04	0.79	0.02	0.78	0.04	0.73	0.08
65	0.96	0.04	0.94	0.04	0.93	0.05	0.81	0.05	0.81	0.06	0.73	0.07	0.78	0.05	0.78	0.04	0.76	0.06
66	0.96	0.04	0.92	0.03	0.93	0.05	0.8	0.07	0.78	0.04	0.72	0.06	0.79	0.04	0.79	0.04	0.74	0.05
67	0.95	0.05	0.92	0.05	0.93	0.05	0.83	0.06	0.79	0.09	0.79	0.07	0.8	0.04	0.76	0.04	0.72	0.07

TABLE S4: RFE SNPS SELECTION – SKYN 56 – Machine learning models (SVM, RF and MLP) average accuracy (avg) and standard deviation (std) of ten-fold cross validation.

SNP	SKYN – 51						SKYN – 52						SKYN – 6					
	White+Pale, Beige, LBrown, MBrown, DBrown						White, Pale, Beige, LBrown, MBrown+DBrown						White, Pale, Beige, LBrown, MBrown, DBrown					
	SVM		RF		MLP		SVM		RF		MLP		SVM		RF		MLP	
	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std
2	0.66	0.03	0.66	0.03	0.66	0.03	0.47	0.03	0.47	0.03	0.47	0.03	0.41	0.03	0.41	0.03	0.41	0.03
3	0.67	0.05	0.68	0.03	0.68	0.03	0.46	0.02	0.48	0.03	0.47	0.03	0.43	0.04	0.43	0.04	0.44	0.04
4	0.67	0.03	0.68	0.03	0.68	0.04	0.45	0.03	0.46	0.04	0.46	0.05	0.43	0.08	0.44	0.09	0.46	0.09
5	0.66	0.03	0.69	0.02	0.69	0.03	0.46	0.03	0.47	0.04	0.5	0.08	0.45	0.06	0.46	0.07	0.47	0.06
6	0.68	0.04	0.68	0.04	0.7	0.03	0.47	0.03	0.47	0.02	0.48	0.06	0.44	0.04	0.45	0.04	0.46	0.08
7	0.68	0.02	0.69	0.03	0.69	0.03	0.46	0.03	0.49	0.06	0.52	0.07	0.45	0.04	0.46	0.04	0.48	0.07
8	0.7	0.04	0.69	0.03	0.67	0.05	0.47	0.04	0.5	0.03	0.48	0.05	0.44	0.06	0.46	0.06	0.47	0.07
9	0.69	0.03	0.67	0.04	0.68	0.03	0.47	0.03	0.47	0.05	0.49	0.07	0.43	0.05	0.45	0.04	0.45	0.07
10	0.69	0.03	0.68	0.03	0.69	0.02	0.47	0.05	0.48	0.05	0.49	0.06	0.45	0.08	0.44	0.07	0.46	0.07
11	0.72	0.04	0.66	0.05	0.68	0.05	0.49	0.03	0.49	0.03	0.48	0.05	0.43	0.04	0.43	0.07	0.48	0.05
12	0.73	0.04	0.74	0.04	0.72	0.05	0.48	0.04	0.47	0.07	0.5	0.06	0.45	0.04	0.46	0.08	0.51	0.05
13	0.73	0.03	0.71	0.06	0.72	0.06	0.51	0.04	0.51	0.02	0.54	0.06	0.48	0.04	0.44	0.05	0.48	0.07
14	0.72	0.07	0.7	0.05	0.73	0.06	0.47	0.07	0.48	0.05	0.51	0.07	0.5	0.06	0.47	0.05	0.51	0.07
15	0.73	0.05	0.74	0.07	0.74	0.05	0.51	0.06	0.48	0.04	0.5	0.05	0.51	0.04	0.5	0.08	0.52	0.1
16	0.72	0.04	0.71	0.04	0.71	0.04	0.51	0.07	0.47	0.06	0.48	0.09	0.48	0.05	0.49	0.07	0.48	0.05
17	0.72	0.04	0.71	0.06	0.71	0.06	0.49	0.05	0.49	0.03	0.52	0.04	0.5	0.06	0.51	0.06	0.51	0.07
18	0.74	0.04	0.71	0.03	0.7	0.05	0.51	0.04	0.48	0.07	0.5	0.05	0.48	0.09	0.51	0.08	0.52	0.08
19	0.71	0.02	0.69	0.04	0.71	0.05	0.43	0.06	0.49	0.1	0.51	0.08	0.49	0.07	0.46	0.05	0.48	0.09
20	0.73	0.04	0.7	0.03	0.67	0.04	0.48	0.06	0.5	0.04	0.47	0.07	0.46	0.07	0.49	0.11	0.48	0.07
21	0.71	0.05	0.69	0.05	0.71	0.06	0.46	0.04	0.48	0.04	0.45	0.08	0.49	0.07	0.47	0.08	0.51	0.07
22	0.7	0.05	0.71	0.05	0.67	0.09	0.48	0.07	0.48	0.07	0.5	0.08	0.47	0.07	0.47	0.05	0.49	0.06
23	0.71	0.06	0.71	0.05	0.68	0.06	0.47	0.02	0.52	0.07	0.52	0.07	0.46	0.06	0.46	0.07	0.47	0.05
24	0.71	0.03	0.73	0.04	0.66	0.08	0.47	0.04	0.52	0.07	0.49	0.09	0.44	0.03	0.45	0.03	0.47	0.06
25	0.71	0.04	0.7	0.04	0.68	0.05	0.48	0.06	0.51	0.08	0.47	0.08	0.49	0.06	0.48	0.07	0.49	0.09
26	0.73	0.04	0.71	0.05	0.7	0.06	0.47	0.07	0.49	0.06	0.48	0.08	0.45	0.04	0.45	0.05	0.45	0.09
27	0.72	0.04	0.71	0.03	0.69	0.07	0.48	0.04	0.47	0.03	0.46	0.08	0.45	0.05	0.45	0.06	0.47	0.08
28	0.72	0.02	0.69	0.05	0.72	0.09	0.5	0.06	0.53	0.09	0.49	0.03	0.46	0.06	0.44	0.06	0.44	0.11
29	0.72	0.04	0.7	0.07	0.66	0.07	0.51	0.08	0.49	0.06	0.48	0.06	0.46	0.07	0.47	0.06	0.48	0.07
30	0.71	0.04	0.69	0.03	0.66	0.07	0.5	0.07	0.53	0.05	0.48	0.1	0.49	0.04	0.48	0.09	0.46	0.05
31	0.73	0.05	0.7	0.05	0.69	0.06	0.51	0.04	0.52	0.08	0.48	0.11	0.5	0.08	0.46	0.07	0.43	0.09
32	0.72	0.06	0.7	0.03	0.68	0.04	0.5	0.07	0.5	0.09	0.46	0.1	0.45	0.08	0.47	0.06	0.42	0.08
33	0.7	0.04	0.7	0.04	0.67	0.07	0.53	0.05	0.51	0.06	0.47	0.09	0.47	0.05	0.48	0.07	0.5	0.08
34	0.73	0.04	0.7	0.04	0.7	0.03	0.48	0.08	0.53	0.06	0.46	0.12	0.49	0.07	0.46	0.1	0.48	0.07
35	0.74	0.04	0.72	0.05	0.72	0.06	0.52	0.05	0.52	0.06	0.46	0.07	0.46	0.09	0.47	0.06	0.44	0.05
36	0.73	0.04	0.73	0.04	0.72	0.05	0.52	0.04	0.55	0.06	0.48	0.05	0.47	0.06	0.45	0.1	0.43	0.06
37	0.75	0.03	0.73	0.03	0.7	0.07	0.53	0.08	0.51	0.07	0.48	0.09	0.47	0.06	0.47	0.06	0.47	0.06
38	0.73	0.06	0.73	0.05	0.7	0.06	0.52	0.08	0.5	0.07	0.48	0.05	0.48	0.1	0.51	0.07	0.46	0.09
39	0.75	0.04	0.71	0.04	0.67	0.08	0.52	0.06	0.5	0.08	0.46	0.08	0.5	0.07	0.46	0.05	0.44	0.08
40	0.74	0.04	0.72	0.06	0.66	0.07	0.53	0.06	0.48	0.08	0.48	0.08	0.5	0.05	0.51	0.04	0.43	0.1
41	0.76	0.02	0.76	0.03	0.74	0.03	0.53	0.05	0.52	0.05	0.49	0.06	0.49	0.07	0.49	0.07	0.44	0.07
42	0.76	0.03	0.75	0.04	0.64	0.07	0.52	0.05	0.48	0.07	0.48	0.05	0.49	0.06	0.48	0.04	0.42	0.06
43	0.74	0.05	0.72	0.03	0.74	0.02	0.52	0.08	0.53	0.05	0.48	0.05	0.48	0.09	0.47	0.11	0.46	0.07
44	0.74	0.04	0.74	0.05	0.7	0.06	0.49	0.07	0.51	0.06	0.52	0.1	0.52	0.07	0.48	0.07	0.42	0.06
45	0.76	0.03	0.72	0.04	0.71	0.05	0.56	0.08	0.53	0.05	0.5	0.07	0.49	0.05	0.49	0.06	0.43	0.08
46	0.75	0.01	0.73	0.02	0.7	0.07	0.51	0.08	0.54	0.05	0.48	0.07	0.49	0.06	0.44	0.08	0.46	0.06
47	0.76	0.05	0.73	0.03	0.7	0.05	0.52	0.07	0.55	0.07	0.48	0.04	0.51	0.07	0.46	0.07	0.42	0.05
48	0.75	0.04	0.72	0.06	0.69	0.07	0.5	0.06	0.51	0.05	0.53	0.03	0.5	0.06	0.45	0.1	0.43	0.07
49	0.74	0.03	0.74	0.05	0.69	0.05	0.53	0.06	0.54	0.06	0.47	0.04	0.47	0.09	0.47	0.07	0.45	0.06
50	0.75	0.03	0.76	0.05	0.67	0.07	0.53	0.07	0.49	0.08	0.42	0.09	0.49	0.06	0.48	0.05	0.46	0.11
51	0.75	0.02	0.76	0.02	0.71	0.06	0.5	0.04	0.53	0.06	0.42	0.08	0.49	0.07	0.48	0.07	0.41	0.08
52	0.73	0.05	0.71	0.04	0.72	0.05	0.51	0.06	0.55	0.05	0.47	0.08	0.48	0.05	0.5	0.06	0.46	0.09
53	0.75	0.04	0.73	0.03	0.7	0.06	0.49	0.06	0.51	0.05	0.42	0.06	0.51	0.07	0.5	0.06	0.42	0.08
54	0.74	0.04	0.74	0.05	0.67	0.05	0.46	0.07	0.47	0.04	0.48	0.07	0.49	0.08	0.48	0.09	0.44	0.06
55	0.75	0.03	0.74	0.03	0.67	0.06	0.5	0.05	0.51	0.06	0.44	0.08	0.49	0.1	0.48	0.08	0.46	0.14
56	0.75	0.04	0.76	0.03	0.7	0.05	0.5	0.07	0.51	0.08	0.44	0.07	0.51	0.09	0.5	0.08	0.44	0.13
57	0.75	0.04	0.74	0.04	0.67	0.07	0.47	0.09	0.51	0.09	0.47	0.07	0.5	0.08	0.48	0.08	0.45	0.08
58	0.77	0.03	0.78	0.03	0.7	0.04	0.52	0.07	0.52	0.06	0.5	0.1	0.49	0.08	0.46	0.08	0.44	0.05
59	0.75	0.03	0.74	0.03	0.68	0.06	0.53	0.06	0.48	0.09	0.49	0.05	0.5	0.06	0.52	0.09	0.42	0.08
60	0.76	0.04	0.74	0.03	0.67	0.07	0.5	0.08	0.51	0.08	0.41	0.1	0.48	0.04	0.49	0.05	0.42	0.1
61	0.75	0.04	0.73	0.04	0.7	0.04	0.49	0.06	0.5	0.08	0.46	0.09	0.44	0.08	0.49	0.07	0.44	0.1
62	0.75	0.03	0.75	0.04	0.69	0.05	0.51	0.09	0.53	0.09	0.42	0.08	0.48	0.07	0.47	0.06	0.44	0.1
63	0.76	0.05	0.76	0.04	0.66	0.06	0.49	0.03	0.48	0.08	0.45	0.07	0.44	0.06	0.45	0.08	0.41	0.06
64	0.77	0.04	0.75	0.06	0.69	0.09	0.53	0.06	0.51	0.07	0.44	0.04	0.44	0.06	0.45	0.07	0.42	0.06
65	0.74	0.04	0.73	0.03	0.67	0.05	0.48	0.08	0.52	0.11	0.49	0.07	0.43	0.09	0.49	0.09	0.38	0.07
66	0.74	0.04	0.73	0.03	0.69	0.06	0.5	0.05	0.54	0.07	0.44	0.08	0.42	0.1	0.45	0.06	0.41	0.09
67	0.75	0.06	0.73	0.05	0.7	0.05	0.49	0.08	0.54	0.06	0.45	0.09	0.46	0.08	0.49	0.09	0.43	0.12

TABLE S5: X^2 SNPS SELECTION – SKIN 234 – Machine learning models (SVM, RF and MLP) average accuracy (avg) and standard deviation (std) of ten-fold cross validation.

SNP	SKYN – 2						SKYN – 3						SKYN – 4					
	White+Pale+Beige, LBrown+MBrown+DBrown						White+Pale, Beige+LBrown, MBrown+DBrown						White+Pale, Beige, LBrown, MBrown+DBrown					
	SVM		RF		MLP		SVM		RF		MLP		SVM		RF		MLP	
	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std
2	0,9	0,04	0,91	0,04	0,91	0,03	0,74	0,03	0,73	0,03	0,73	0,03	0,74	0,03	0,74	0,06	0,73	0,04
3	0,9	0,03	0,91	0,03	0,92	0,02	0,74	0,04	0,75	0,04	0,76	0,04	0,73	0,03	0,74	0,03	0,75	0,03
4	0,91	0,04	0,91	0,03	0,93	0,03	0,73	0,04	0,75	0,03	0,75	0,05	0,74	0,04	0,74	0,03	0,74	0,04
5	0,92	0,04	0,92	0,05	0,92	0,05	0,75	0,03	0,74	0,06	0,75	0,05	0,76	0,03	0,76	0,02	0,75	0,05
6	0,93	0,03	0,91	0,04	0,92	0,04	0,75	0,03	0,76	0,02	0,75	0,06	0,76	0,04	0,76	0,02	0,75	0,03
7	0,95	0,02	0,93	0,02	0,95	0,02	0,77	0,02	0,77	0,02	0,78	0,06	0,77	0,05	0,76	0,03	0,76	0,05
8	0,93	0,04	0,93	0,05	0,94	0,03	0,77	0,02	0,77	0,03	0,77	0,05	0,78	0,04	0,79	0,02	0,77	0,06
9	0,95	0,02	0,94	0,02	0,95	0,03	0,77	0,03	0,76	0,03	0,78	0,05	0,8	0,03	0,78	0,03	0,77	0,03
10	0,96	0,02	0,94	0,04	0,95	0,03	0,76	0,02	0,77	0,02	0,75	0,05	0,79	0,04	0,78	0,02	0,78	0,02
11	0,95	0,02	0,94	0,04	0,95	0,03	0,76	0,02	0,77	0,02	0,79	0,04	0,79	0,03	0,79	0,03	0,77	0,04
12	0,96	0,01	0,95	0,02	0,92	0,03	0,77	0,03	0,77	0,03	0,78	0,06	0,79	0,03	0,77	0,03	0,79	0,04
13	0,95	0,02	0,94	0,03	0,96	0,02	0,77	0,02	0,76	0,03	0,77	0,05	0,78	0,04	0,77	0,02	0,78	0,03
14	0,95	0,03	0,93	0,04	0,95	0,04	0,77	0,05	0,76	0,04	0,76	0,06	0,79	0,02	0,78	0,02	0,78	0,05
15	0,96	0,03	0,94	0,04	0,94	0,05	0,76	0,03	0,77	0,04	0,76	0,07	0,78	0,05	0,78	0,04	0,77	0,06
16	0,95	0,03	0,94	0,02	0,96	0,02	0,77	0,02	0,77	0,02	0,76	0,05	0,78	0,02	0,77	0,02	0,77	0,03
17	0,94	0,05	0,94	0,04	0,93	0,05	0,77	0,03	0,77	0,03	0,74	0,07	0,79	0,04	0,76	0,04	0,75	0,05
18	0,95	0,04	0,94	0,03	0,93	0,04	0,77	0,04	0,78	0,03	0,75	0,07	0,8	0,03	0,76	0,05	0,77	0,03
19	0,95	0,05	0,94	0,04	0,93	0,05	0,77	0,04	0,77	0,03	0,74	0,06	0,79	0,03	0,77	0,02	0,74	0,06
20	0,95	0,04	0,95	0,03	0,91	0,05	0,77	0,02	0,76	0,02	0,76	0,08	0,79	0,03	0,77	0,03	0,77	0,03
21	0,96	0,03	0,93	0,05	0,91	0,07	0,77	0,03	0,77	0,04	0,77	0,07	0,78	0,04	0,76	0,02	0,78	0,05
22	0,96	0,02	0,94	0,02	0,93	0,05	0,78	0,03	0,77	0,02	0,75	0,04	0,79	0,03	0,77	0,02	0,73	0,04
23	0,96	0,02	0,93	0,03	0,93	0,04	0,76	0,02	0,76	0,02	0,74	0,05	0,78	0,03	0,77	0,03	0,76	0,05
24	0,95	0,03	0,95	0,03	0,91	0,06	0,76	0,04	0,74	0,05	0,73	0,04	0,77	0,03	0,76	0,03	0,73	0,05
25	0,95	0,04	0,94	0,03	0,94	0,06	0,78	0,04	0,77	0,03	0,73	0,07	0,79	0,04	0,78	0,03	0,76	0,05
26	0,97	0,04	0,95	0,04	0,94	0,04	0,78	0,06	0,77	0,04	0,71	0,06	0,79	0,03	0,77	0,04	0,69	0,07
27	0,97	0,03	0,95	0,02	0,94	0,05	0,78	0,05	0,76	0,07	0,71	0,08	0,8	0,04	0,79	0,04	0,74	0,07
28	0,97	0,03	0,94	0,04	0,94	0,04	0,77	0,03	0,75	0,04	0,69	0,07	0,81	0,03	0,79	0,03	0,72	0,06
29	0,96	0,03	0,95	0,03	0,94	0,02	0,79	0,04	0,75	0,05	0,69	0,06	0,79	0,03	0,8	0,03	0,73	0,08
30	0,96	0,03	0,93	0,05	0,94	0,04	0,79	0,05	0,77	0,04	0,73	0,07	0,79	0,03	0,78	0,05	0,74	0,04
31	0,96	0,02	0,93	0,03	0,93	0,02	0,8	0,07	0,8	0,04	0,74	0,06	0,79	0,02	0,76	0,04	0,72	0,07
32	0,96	0,03	0,94	0,03	0,93	0,05	0,8	0,05	0,79	0,03	0,77	0,06	0,79	0,05	0,77	0,05	0,72	0,04
33	0,96	0,03	0,95	0,03	0,95	0,04	0,78	0,06	0,78	0,04	0,74	0,05	0,78	0,04	0,76	0,04	0,73	0,05
34	0,97	0,03	0,95	0,03	0,96	0,02	0,8	0,04	0,8	0,03	0,77	0,08	0,79	0,04	0,8	0,01	0,76	0,06
35	0,97	0,02	0,94	0,04	0,93	0,03	0,79	0,05	0,77	0,07	0,72	0,1	0,78	0,04	0,76	0,04	0,71	0,06
36	0,96	0,04	0,95	0,03	0,92	0,03	0,78	0,04	0,8	0,04	0,72	0,1	0,78	0,04	0,76	0,05	0,69	0,08
37	0,96	0,04	0,95	0,05	0,93	0,05	0,81	0,03	0,8	0,04	0,75	0,07	0,79	0,04	0,77	0,04	0,73	0,07
38	0,95	0,05	0,94	0,04	0,94	0,05	0,79	0,05	0,79	0,06	0,78	0,08	0,78	0,04	0,77	0,04	0,7	0,09
39	0,96	0,02	0,95	0,03	0,89	0,06	0,77	0,05	0,78	0,04	0,74	0,08	0,8	0,04	0,81	0,04	0,74	0,05
40	0,95	0,03	0,94	0,04	0,93	0,05	0,78	0,07	0,78	0,07	0,72	0,06	0,78	0,03	0,78	0,05	0,71	0,11
41	0,93	0,05	0,93	0,04	0,9	0,07	0,79	0,04	0,79	0,03	0,74	0,08	0,78	0,03	0,74	0,04	0,69	0,06
42	0,96	0,03	0,93	0,04	0,93	0,05	0,78	0,05	0,8	0,06	0,73	0,08	0,79	0,06	0,76	0,04	0,76	0,08
43	0,96	0,03	0,94	0,04	0,95	0,03	0,8	0,04	0,78	0,04	0,74	0,06	0,79	0,05	0,76	0,06	0,76	0,05
44	0,95	0,04	0,93	0,05	0,94	0,05	0,8	0,05	0,8	0,07	0,76	0,08	0,78	0,04	0,76	0,07	0,72	0,04
45	0,95	0,04	0,94	0,04	0,93	0,05	0,8	0,07	0,81	0,08	0,73	0,08	0,79	0,07	0,77	0,08	0,71	0,04
46	0,95	0,03	0,93	0,03	0,93	0,04	0,79	0,06	0,81	0,06	0,72	0,05	0,78	0,07	0,76	0,06	0,73	0,09
47	0,96	0,04	0,95	0,03	0,92	0,04	0,8	0,05	0,8	0,05	0,77	0,07	0,78	0,04	0,77	0,07	0,71	0,05
48	0,96	0,03	0,92	0,06	0,93	0,02	0,8	0,07	0,8	0,05	0,7	0,08	0,78	0,05	0,78	0,03	0,74	0,07
49	0,95	0,03	0,95	0,04	0,94	0,04	0,8	0,07	0,79	0,05	0,75	0,06	0,8	0,05	0,79	0,05	0,68	0,11
50	0,95	0,03	0,95	0,03	0,95	0,03	0,79	0,05	0,8	0,06	0,74	0,11	0,79	0,05	0,77	0,06	0,74	0,05
51	0,95	0,05	0,94	0,06	0,93	0,06	0,79	0,04	0,79	0,06	0,72	0,11	0,79	0,05	0,77	0,03	0,73	0,05
52	0,95	0,05	0,93	0,04	0,93	0,04	0,8	0,03	0,8	0,05	0,72	0,08	0,79	0,06	0,78	0,03	0,7	0,09
53	0,97	0,04	0,95	0,05	0,94	0,03	0,78	0,07	0,79	0,07	0,74	0,06	0,79	0,05	0,78	0,04	0,78	0,07
54	0,97	0,04	0,94	0,05	0,95	0,04	0,8	0,03	0,83	0,06	0,71	0,06	0,78	0,03	0,78	0,03	0,73	0,08
55	0,96	0,04	0,93	0,05	0,96	0,04	0,78	0,06	0,82	0,06	0,75	0,1	0,8	0,04	0,77	0,02	0,72	0,11
56	0,96	0,05	0,93	0,05	0,95	0,05	0,79	0,07	0,81	0,07	0,77	0,08	0,79	0,04	0,77	0,04	0,77	0,09
57	0,96	0,05	0,93	0,05	0,94	0,05	0,8	0,07	0,81	0,06	0,77	0,07	0,78	0,04	0,78	0,04	0,74	0,07
58	0,96	0,04	0,93	0,03	0,95	0,04	0,78	0,07	0,78	0,05	0,76	0,07	0,77	0,05	0,78	0,06	0,75	0,05
59	0,95	0,04	0,93	0,04	0,94	0,05	0,78	0,04	0,81	0,05	0,75	0,07	0,8	0,04	0,77	0,03	0,71	0,06
60	0,96	0,04	0,93	0,03	0,94	0,04	0,79	0,06	0,81	0,06	0,76	0,06	0,78	0,03	0,76	0,05	0,74	0,07
61	0,97	0,03	0,92	0,04	0,94	0,02	0,8	0,06	0,81	0,06	0,72	0,05	0,79	0,04	0,79	0,04	0,71	0,08
62	0,96	0,03	0,95	0,03	0,96	0,02	0,79	0,05	0,79	0,06	0,74	0,08	0,79	0,04	0,78	0,05	0,74	0,09
63	0,96	0,03	0,93	0,04	0,93	0,06	0,81	0,05	0,79	0,05	0,76	0,08	0,79	0,03	0,76	0,03	0,72	0,06
64	0,95	0,04	0,93	0,04	0,93	0,04	0,79	0,06	0,79	0,07	0,77	0,06	0,78	0,04	0,79	0,05	0,74	0,06
65	0,95	0,04	0,93	0,04	0,93	0,04	0,81	0,06	0,81	0,08	0,74	0,06	0,79	0,04	0,79	0,04	0,73	0,09
66	0,96	0,03	0,93	0,04	0,93	0,03	0,8	0,05	0,78	0,06	0,75	0,07	0,8	0,04	0,77	0,05	0,75	0,07
67	0,95	0,04	0,94	0,06	0,94	0,04	0,83	0,07	0,77	0,08	0,77	0,08	0,79	0,04	0,77	0,04	0,74	0,04

TABLE S6: X^2 SNPS SELECTION – SKIN 56 – Machine learning models (SVM, RF and MLP) average accuracy (avg) and standard deviation (std) of ten-fold cross validation.

SNP	SKYN – 51						SKYN – 52						SKYN – 6					
	White+Pale, Beige, LBrown, MBrown, DBrown						White, Pale, Beige, LBrown, MBrown+DBrown						White, Pale, Beige, LBrown, MBrown, DBrown					
	SVM		RF		MLP		SVM		RF		MLP		SVM		RF		MLP	
	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std
2	0,66	0,03	0,65	0,03	0,65	0,03	0,48	0,02	0,48	0,02	0,49	0,03	0,41	0,03	0,41	0,03	0,39	0,04
3	0,66	0,03	0,66	0,03	0,64	0,05	0,48	0,04	0,5	0,03	0,5	0,04	0,41	0,03	0,4	0,03	0,39	0,04
4	0,69	0,04	0,69	0,03	0,71	0,05	0,48	0,04	0,49	0,02	0,5	0,04	0,44	0,04	0,44	0,04	0,44	0,06
5	0,72	0,02	0,71	0,04	0,67	0,05	0,49	0,03	0,51	0,03	0,5	0,06	0,45	0,02	0,45	0,03	0,47	0,05
6	0,71	0,03	0,7	0,02	0,71	0,04	0,51	0,03	0,5	0,02	0,49	0,06	0,46	0,04	0,45	0,02	0,45	0,04
7	0,7	0,05	0,72	0,03	0,71	0,03	0,49	0,04	0,5	0,02	0,52	0,03	0,45	0,04	0,44	0,03	0,45	0,04
8	0,73	0,04	0,71	0,05	0,75	0,04	0,52	0,04	0,51	0,01	0,51	0,05	0,46	0,07	0,45	0,04	0,45	0,06
9	0,73	0,03	0,73	0,02	0,72	0,07	0,51	0,03	0,5	0,04	0,54	0,07	0,48	0,08	0,48	0,07	0,51	0,06
10	0,74	0,04	0,73	0,04	0,74	0,03	0,51	0,06	0,54	0,05	0,55	0,06	0,51	0,06	0,45	0,04	0,47	0,04
11	0,74	0,04	0,72	0,05	0,74	0,05	0,51	0,04	0,55	0,05	0,53	0,09	0,51	0,07	0,47	0,06	0,46	0,07
12	0,75	0,05	0,72	0,04	0,68	0,07	0,53	0,05	0,55	0,06	0,55	0,08	0,51	0,08	0,48	0,07	0,51	0,08
13	0,75	0,05	0,7	0,03	0,72	0,04	0,53	0,05	0,51	0,03	0,54	0,06	0,51	0,07	0,47	0,08	0,48	0,06
14	0,75	0,04	0,73	0,04	0,72	0,07	0,52	0,04	0,52	0,04	0,5	0,02	0,49	0,07	0,48	0,08	0,44	0,08
15	0,74	0,04	0,71	0,04	0,71	0,05	0,54	0,06	0,52	0,06	0,52	0,07	0,51	0,06	0,47	0,06	0,45	0,07
16	0,75	0,03	0,71	0,05	0,74	0,06	0,5	0,06	0,53	0,07	0,55	0,05	0,5	0,06	0,46	0,06	0,45	0,08
17	0,75	0,04	0,73	0,04	0,71	0,05	0,51	0,04	0,54	0,05	0,47	0,06	0,52	0,06	0,46	0,08	0,45	0,05
18	0,74	0,05	0,72	0,03	0,73	0,04	0,53	0,06	0,52	0,05	0,53	0,07	0,51	0,07	0,45	0,07	0,46	0,08
19	0,73	0,04	0,71	0,05	0,69	0,05	0,52	0,04	0,53	0,07	0,51	0,07	0,5	0,07	0,48	0,08	0,48	0,06
20	0,74	0,03	0,71	0,04	0,7	0,06	0,54	0,04	0,49	0,08	0,52	0,09	0,5	0,08	0,47	0,07	0,47	0,1
21	0,73	0,05	0,7	0,05	0,7	0,04	0,51	0,08	0,5	0,08	0,51	0,08	0,48	0,05	0,48	0,04	0,51	0,05
22	0,74	0,03	0,72	0,04	0,7	0,05	0,5	0,05	0,49	0,06	0,49	0,06	0,46	0,07	0,44	0,06	0,45	0,07
23	0,75	0,03	0,72	0,03	0,7	0,05	0,52	0,04	0,53	0,08	0,55	0,07	0,49	0,03	0,46	0,05	0,45	0,08
24	0,75	0,03	0,72	0,02	0,7	0,05	0,5	0,05	0,48	0,07	0,49	0,08	0,45	0,06	0,47	0,09	0,45	0,12
25	0,75	0,04	0,73	0,04	0,72	0,05	0,5	0,05	0,53	0,07	0,47	0,06	0,48	0,09	0,45	0,06	0,46	0,09
26	0,76	0,05	0,74	0,03	0,75	0,05	0,48	0,06	0,5	0,07	0,45	0,07	0,47	0,08	0,48	0,05	0,42	0,06
27	0,77	0,05	0,75	0,04	0,69	0,07	0,47	0,06	0,48	0,07	0,47	0,06	0,46	0,04	0,49	0,08	0,43	0,07
28	0,76	0,04	0,73	0,05	0,67	0,08	0,48	0,05	0,52	0,07	0,51	0,09	0,48	0,1	0,46	0,04	0,43	0,06
29	0,76	0,05	0,75	0,03	0,71	0,05	0,51	0,04	0,51	0,07	0,5	0,08	0,51	0,08	0,51	0,05	0,39	0,09
30	0,74	0,03	0,75	0,03	0,7	0,05	0,56	0,06	0,52	0,05	0,49	0,08	0,47	0,06	0,47	0,04	0,45	0,06
31	0,75	0,03	0,74	0,05	0,72	0,1	0,51	0,08	0,5	0,06	0,47	0,05	0,51	0,06	0,46	0,09	0,42	0,11
32	0,75	0,03	0,76	0,04	0,67	0,05	0,5	0,04	0,51	0,08	0,48	0,07	0,5	0,04	0,46	0,07	0,47	0,07
33	0,75	0,03	0,75	0,04	0,7	0,09	0,51	0,04	0,51	0,05	0,43	0,1	0,51	0,07	0,5	0,08	0,42	0,11
34	0,74	0,03	0,75	0,03	0,7	0,07	0,54	0,07	0,49	0,06	0,48	0,06	0,51	0,06	0,51	0,08	0,46	0,08
35	0,76	0,04	0,73	0,05	0,69	0,08	0,51	0,04	0,52	0,06	0,44	0,09	0,48	0,09	0,52	0,05	0,43	0,08
36	0,74	0,03	0,75	0,04	0,69	0,06	0,51	0,06	0,51	0,06	0,51	0,08	0,46	0,08	0,49	0,06	0,43	0,05
37	0,74	0,04	0,75	0,04	0,7	0,05	0,52	0,09	0,51	0,07	0,47	0,11	0,44	0,06	0,47	0,05	0,44	0,09
38	0,73	0,04	0,74	0,04	0,72	0,06	0,52	0,1	0,57	0,09	0,46	0,07	0,48	0,07	0,5	0,07	0,43	0,07
39	0,75	0,04	0,75	0,04	0,69	0,07	0,52	0,07	0,52	0,07	0,46	0,07	0,45	0,05	0,48	0,07	0,44	0,07
40	0,75	0,05	0,73	0,06	0,72	0,05	0,52	0,07	0,49	0,05	0,5	0,09	0,46	0,05	0,48	0,07	0,4	0,06
41	0,75	0,04	0,74	0,03	0,72	0,07	0,53	0,03	0,53	0,07	0,46	0,06	0,5	0,05	0,46	0,09	0,49	0,07
42	0,75	0,03	0,74	0,06	0,68	0,1	0,52	0,07	0,49	0,08	0,49	0,07	0,48	0,07	0,5	0,06	0,42	0,08
43	0,76	0,04	0,72	0,05	0,71	0,05	0,52	0,07	0,46	0,11	0,47	0,07	0,46	0,09	0,49	0,1	0,41	0,08
44	0,75	0,05	0,73	0,05	0,74	0,07	0,51	0,06	0,53	0,05	0,5	0,08	0,51	0,05	0,51	0,08	0,44	0,1
45	0,76	0,05	0,75	0,06	0,72	0,05	0,55	0,09	0,55	0,06	0,46	0,07	0,47	0,05	0,45	0,07	0,41	0,08
46	0,75	0,04	0,73	0,05	0,67	0,07	0,54	0,09	0,52	0,09	0,44	0,09	0,52	0,1	0,44	0,09	0,45	0,1
47	0,74	0,06	0,73	0,05	0,69	0,06	0,51	0,06	0,52	0,08	0,49	0,06	0,51	0,08	0,48	0,07	0,48	0,09
48	0,73	0,04	0,74	0,05	0,71	0,06	0,53	0,04	0,56	0,07	0,49	0,09	0,49	0,08	0,49	0,06	0,49	0,05
49	0,74	0,05	0,74	0,04	0,67	0,07	0,53	0,09	0,53	0,09	0,42	0,1	0,5	0,1	0,52	0,07	0,42	0,06
50	0,74	0,03	0,74	0,05	0,7	0,04	0,53	0,09	0,53	0,11	0,49	0,1	0,48	0,07	0,46	0,08	0,44	0,05
51	0,75	0,04	0,71	0,04	0,72	0,06	0,55	0,05	0,52	0,07	0,45	0,09	0,48	0,07	0,47	0,1	0,46	0,15
52	0,75	0,04	0,74	0,05	0,69	0,08	0,54	0,08	0,53	0,1	0,48	0,11	0,47	0,06	0,5	0,1	0,46	0,05
53	0,74	0,04	0,74	0,03	0,69	0,06	0,48	0,08	0,52	0,05	0,52	0,06	0,47	0,07	0,46	0,1	0,44	0,05
54	0,75	0,04	0,73	0,04	0,69	0,07	0,52	0,07	0,47	0,06	0,46	0,04	0,48	0,08	0,52	0,07	0,37	0,08
55	0,75	0,02	0,74	0,05	0,71	0,06	0,48	0,07	0,52	0,07	0,46	0,11	0,46	0,1	0,5	0,05	0,46	0,05
56	0,74	0,06	0,74	0,05	0,7	0,06	0,54	0,08	0,5	0,09	0,46	0,07	0,41	0,11	0,47	0,12	0,45	0,1
57	0,73	0,04	0,74	0,03	0,68	0,07	0,48	0,04	0,5	0,06	0,47	0,09	0,49	0,1	0,47	0,1	0,4	0,08
58	0,74	0,03	0,71	0,03	0,69	0,07	0,52	0,08	0,49	0,08	0,46	0,08	0,45	0,08	0,47	0,12	0,45	0,07
59	0,74	0,05	0,75	0,04	0,7	0,09	0,51	0,09	0,53	0,07	0,46	0,06	0,45	0,04	0,44	0,07	0,41	0,07
60	0,74	0,04	0,71	0,03	0,69	0,05	0,49	0,06	0,52	0,09	0,47	0,08	0,46	0,06	0,44	0,06	0,4	0,09
61	0,73	0,03	0,74	0,04	0,71	0,09	0,47	0,07	0,51	0,1	0,48	0,05	0,5	0,06	0,51	0,09	0,43	0,07
62	0,76	0,06	0,74	0,03	0,7	0,07	0,51	0,11	0,56	0,1	0,49	0,08	0,46	0,08	0,5	0,11	0,39	0,1
63	0,74	0,05	0,73	0,06	0,72	0,03	0,48	0,04	0,49	0,08	0,45	0,06	0,45	0,07	0,45	0,1	0,42	0,12
64	0,75	0,05	0,74	0,05	0,69	0,08	0,5	0,07	0,54	0,08	0,45	0,07	0,46	0,09	0,47	0,09	0,42	0,08
65	0,75	0,05	0,78	0,05	0,72	0,07	0,5	0,07	0,48	0,1	0,43	0,08	0,46	0,1	0,46	0,08	0,4	0,05
66	0,75	0,05	0,74	0,05	0,68	0,08	0,49	0,06	0,54	0,06	0,44	0,11	0,46	0,06	0,51	0,07	0,42	0,08
67	0,75	0,05	0,75	0,04	0,7	0,09	0,48	0,04	0,5	0,08	0,44	0,07	0,43	0,08	0,45	0,06	0,41	0,09

	'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']
--	---	---

TABLE S8: SNP SELECTION BY RFE AND X² – EYE 3.2 CLASSES

SNPs	RFE	X2
2	['rs1129038', 'rs1448484']	['rs1129038', 'rs12913832']
3	['rs11230664', 'rs1129038', 'rs1448484']	['rs1129038', 'rs12913832', 'rs1426654']
4	['rs11230664', 'rs1129038', 'rs12913832', 'rs1448484']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1426654']
5	['rs11230664', 'rs1129038', 'rs12913832', 'rs1448484', 'rs2402130']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654']
6	['rs11230664', 'rs1129038', 'rs12913832', 'rs1448484', 'rs2402130', 'rs3794606']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654', 'rs895829']
7	['rs11230664', 'rs1129038', 'rs12913832', 'rs1448484', 'rs2402130', 'rs3794606', 'rs895829']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654', 'rs4778232', 'rs895829']
8	['rs11230664', 'rs1129038', 'rs12913832', 'rs1426654', 'rs1448484', 'rs2402130', 'rs3794606', 'rs895829']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654', 'rs3794606', 'rs4778232', 'rs895829']
9	['rs11230664', 'rs1129038', 'rs12913832', 'rs1426654', 'rs1448484', 'rs1800407', 'rs2402130', 'rs3794606', 'rs895829']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1597196', 'rs3794606', 'rs4778232', 'rs895829']
10	['rs1037208', 'rs11230664', 'rs1129038', 'rs12913832', 'rs1426654', 'rs1448484', 'rs1800407', 'rs2402130', 'rs3794606', 'rs895829']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1597196', 'rs3794606', 'rs4778232', 'rs8039195', 'rs895829']
11	['rs1037208', 'rs11230664', 'rs1129038', 'rs12913832', 'rs1426654', 'rs1448484', 'rs1800407', 'rs2402130', 'rs3794606', 'rs6497271', 'rs895829']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1597196', 'rs2238289', 'rs3794606', 'rs4778232', 'rs8039195', 'rs895829']
12	['rs1037208', 'rs11230664', 'rs1129038', 'rs12913832', 'rs1426654', 'rs1448484', 'rs1800407', 'rs2402130', 'rs3794606', 'rs6497271', 'rs6510760', 'rs895829']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1448484', 'rs1597196', 'rs2238289', 'rs3794606', 'rs4778232', 'rs8039195', 'rs895829']
13	['rs1037208', 'rs11230664', 'rs1129038', 'rs12913832', 'rs1426654', 'rs1448484', 'rs1800407', 'rs2402130', 'rs3794606', 'rs4778232', 'rs6497271', 'rs6510760', 'rs895829']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1448484', 'rs1597196', 'rs2238289', 'rs2240203', 'rs3794606', 'rs4778232', 'rs8039195', 'rs895829']
14	['rs1037208', 'rs10424065', 'rs11230664', 'rs1129038', 'rs12913832', 'rs1426654', 'rs1448484', 'rs1800407', 'rs2402130', 'rs3794606', 'rs4778232', 'rs6497271', 'rs6510760', 'rs895829']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16950987', 'rs2238289', 'rs2240203', 'rs3794606', 'rs4778232', 'rs8039195', 'rs895829']
15	['rs1037208', 'rs10424065', 'rs11230664', 'rs1129038', 'rs12913832', 'rs1426654', 'rs1448484', 'rs1800407', 'rs1805006', 'rs2402130', 'rs3794606', 'rs4778232', 'rs6497271', 'rs6510760', 'rs895829']	['rs11230664', 'rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16950987', 'rs2238289', 'rs2240203', 'rs3794606', 'rs4778232', 'rs8039195', 'rs895829']

	'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']
67	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']

TABLE S9: SNP SELECTION BY RFE AND X² – EYE 3.3 CLASSES

SNPs	RFE	X2
2	['rs1129038', 'rs1448484']	['rs1129038', 'rs12913832']
3	['rs1129038', 'rs1448484', 'rs6119471']	['rs1129038', 'rs12913832', 'rs1426654']
4	['rs1129038', 'rs12913832', 'rs1448484', 'rs6119471']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1426654']
5	['rs1129038', 'rs12913832', 'rs1448484', 'rs6119471', 'rs7948623']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654']
6	['rs11230664', 'rs1129038', 'rs12913832', 'rs1448484', 'rs6119471', 'rs7948623']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654', 'rs895829']
7	['rs11230664', 'rs1129038', 'rs12913832', 'rs1448484', 'rs1800407', 'rs6119471', 'rs7948623']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654', 'rs16950987', 'rs895829']
8	['rs11230664', 'rs1129038', 'rs12203592', 'rs12913832', 'rs1448484', 'rs1800407', 'rs6119471', 'rs7948623']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654', 'rs16950987', 'rs2240203', 'rs895829']
9	['rs11230664', 'rs1129038', 'rs12203592', 'rs12913832', 'rs1426654', 'rs1448484', 'rs1800407', 'rs6119471', 'rs7948623']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654', 'rs16950987', 'rs2240203', 'rs8039195', 'rs895829']
10	['rs1110400', 'rs11230664', 'rs1129038', 'rs12203592', 'rs12913832', 'rs1426654', 'rs1448484', 'rs1800407', 'rs6119471', 'rs7948623']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1375164', 'rs1426654', 'rs16950987', 'rs2240203', 'rs3794606', 'rs8039195', 'rs895829']

	'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']
66	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']
67	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']

TABLE S10: SNP SELECTION BY RFE AND X² – EYE 4 CLASSES

SNPs	RFE	X2
2	['rs11230664', 'rs1129038']	['rs1129038', 'rs12913832']
3	['rs11230664', 'rs1129038', 'rs6497271']	['rs1129038', 'rs11636232', 'rs12913832']
4	['rs11230664', 'rs1129038', 'rs12913832', 'rs6497271']	['rs1129038', 'rs11636232', 'rs12913832', 'rs1426654']

[illegible]

	'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']
--	---	---

TABLE S12: SNP SELECTION BY RFE AND X² – SKIN 2 CLASSES

SNPs	RFE	X2
2	['rs11230664', 'rs1805006']	['rs11230664', 'rs1426654']
3	['rs11230664', 'rs1805005', 'rs1805006']	['rs11230664', 'rs1426654', 'rs6119471']
4	['rs11230664', 'rs1805005', 'rs1805006', 'rs6497271']	['rs10777129', 'rs11230664', 'rs1426654', 'rs6119471']
5	['rs1042602', 'rs11230664', 'rs1805005', 'rs1805006', 'rs6497271']	['rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471']
6	['rs10424065', 'rs1042602', 'rs11230664', 'rs1805005', 'rs1805006', 'rs6497271']	['rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271']
7	['rs10424065', 'rs1042602', 'rs11230664', 'rs1805005', 'rs1805006', 'rs2424984', 'rs6497271']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271']
8	['rs10424065', 'rs1042602', 'rs11230664', 'rs12913832', 'rs1805005', 'rs1805006', 'rs2424984', 'rs6497271']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471', 'rs642742', 'rs6497271']
9	['rs10424065', 'rs1042602', 'rs10777129', 'rs11230664', 'rs12913832', 'rs1805005', 'rs1805006', 'rs2424984', 'rs6497271']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195']
10	['rs10424065', 'rs1042602', 'rs10777129', 'rs11230664', 'rs12913832', 'rs1805005', 'rs1805006', 'rs2424984', 'rs3768056', 'rs6497271']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs2424984', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195']
11	['rs10424065', 'rs1042602', 'rs10777129', 'rs11230664', 'rs12913832', 'rs1805005', 'rs1805006', 'rs2424984', 'rs3768056', 'rs6497271', 'rs7948623']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1375164', 'rs1426654', 'rs1448484', 'rs2424984', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195']
12	['rs10424065', 'rs1042602', 'rs10777129', 'rs11230664', 'rs12913832', 'rs1800407', 'rs1805005', 'rs1805006', 'rs2424984', 'rs3768056', 'rs6497271', 'rs7948623']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1375164', 'rs1426654', 'rs1448484', 'rs1597196', 'rs2424984', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195']
13	['rs10424065', 'rs1042602', 'rs10777129', 'rs11230664', 'rs12913832', 'rs1426654', 'rs1800407', 'rs1805005', 'rs1805006', 'rs2424984', 'rs3768056', 'rs6497271', 'rs7948623']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1375164', 'rs1426654', 'rs1448484', 'rs1597196', 'rs2424984', 'rs3794606', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195']
14	['rs10424065', 'rs1042602', 'rs10777129', 'rs11230664', 'rs12913832', 'rs1393350', 'rs1426654', 'rs1800407', 'rs1805005', 'rs1805006', 'rs2424984', 'rs3768056', 'rs6497271', 'rs7948623']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1375164', 'rs1426654', 'rs1448484', 'rs1597196', 'rs2424984', 'rs3794606', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195', 'rs895829']
15	['rs10424065', 'rs1042602', 'rs10777129', 'rs11230664', 'rs12913832', 'rs1393350', 'rs1426654', 'rs1800407', 'rs1805005', 'rs1805006', 'rs2424984', 'rs3768056', 'rs6497271', 'rs7948623', 'rs9894429']	['rs10424065', 'rs10777129', 'rs11230664', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1448484', 'rs1597196', 'rs2424984', 'rs3794606', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195', 'rs895829']

66	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs916977', 'rs9894429']
67	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']

TABLE S13: SNP SELECTION BY RFE AND X² – SKIN 3 CLASSES

SNPs	RFE	X2
2	['rs11230664', 'rs1805006']	['rs11230664', 'rs1426654']
3	['rs11230664', 'rs1129038', 'rs1805006']	['rs11230664', 'rs1426654', 'rs1448484']
4	['rs11230664', 'rs1129038', 'rs1805006', 'rs6497271']	['rs11230664', 'rs1426654', 'rs1448484', 'rs6119471']
5	['rs1042602', 'rs11230664', 'rs1129038', 'rs1805006', 'rs6497271']	['rs11230664', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271']
6	['rs1042602', 'rs11230664', 'rs1129038', 'rs1805006', 'rs6497271', 'rs7948623']	['rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271']
7	['rs10424065', 'rs1042602', 'rs11230664', 'rs1129038', 'rs1805006', 'rs6497271', 'rs7948623']	['rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271', 'rs8039195']
8	['rs10424065', 'rs1042602', 'rs11230664', 'rs1129038', 'rs1805006', 'rs1805009', 'rs6497271', 'rs7948623']	['rs10777129', 'rs11230664', 'rs1375164', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271', 'rs8039195']
9	['rs10424065', 'rs1042602', 'rs11230664', 'rs1129038', 'rs1805006', 'rs1805009', 'rs2424984', 'rs6497271', 'rs7948623']	['rs10777129', 'rs11230664', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271', 'rs8039195']

[illegible]

	'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']
67	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']

TABLE S15: SNP SELECTION BY RFE AND X² – SKIN 5.1 CLASSES

SNPs	RFE	X2
2	['rs11230664', 'rs1805006']	['rs11230664', 'rs6119471']
3	['rs11230664', 'rs1448484', 'rs1805006']	['rs11230664', 'rs1426654', 'rs6119471']
4	['rs11230664', 'rs1129038', 'rs1448484', 'rs1805006']	['rs11230664', 'rs1426654', 'rs1448484', 'rs6119471']
5	['rs11230664', 'rs1129038', 'rs1448484', 'rs1805006', 'rs6510760']	['rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471']
6	['rs1042602', 'rs11230664', 'rs1129038', 'rs1448484', 'rs1805006', 'rs6510760']	['rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271']
7	['rs10424065', 'rs1042602', 'rs11230664', 'rs1129038', 'rs1448484', 'rs1805006', 'rs6510760']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271']
8	['rs10424065', 'rs1042602', 'rs11230664', 'rs1129038', 'rs1448484', 'rs1805006', 'rs6510760', 'rs885479']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271', 'rs8039195']
9	['rs10424065', 'rs1042602', 'rs11230664', 'rs1129038', 'rs12203592', 'rs1448484', 'rs1805006', 'rs6510760', 'rs885479']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195']
10	['rs10424065', 'rs1042602', 'rs11230664', 'rs1129038', 'rs12203592', 'rs1448484', 'rs1805006', 'rs1805009', 'rs6510760', 'rs885479']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1375164', 'rs1426654', 'rs1448484', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195']
11	['rs10424065', 'rs1042602', 'rs11230664', 'rs1129038', 'rs12203592', 'rs1426654', 'rs1448484', 'rs1805006', 'rs1805009', 'rs6510760', 'rs885479']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1375164', 'rs1426654', 'rs1448484', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195', 'rs895829']
12	['rs10424065', 'rs1042602', 'rs11230664', 'rs1129038', 'rs12203592', 'rs1426654', 'rs1448484', 'rs16891982', 'rs1805006', 'rs1805009', 'rs6510760', 'rs885479']	['rs10424065', 'rs10777129', 'rs11230664', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1448484', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195', 'rs895829']
13	['rs10424065', 'rs1042602', 'rs11230664', 'rs1129038', 'rs12203592', 'rs1426654', 'rs1448484', 'rs16891982',	['rs10424065', 'rs10777129', 'rs11230664', 'rs1129038', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1448484', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195', 'rs895829']

	'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']
66	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']
67	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']

TABLE S16: SNP SELECTION BY RFE AND X² – SKIN 5.2 CLASSES

SNPs	RFE	X2
2	['rs1110400', 'rs11230664']	['rs11230664', 'rs1426654']
3	['rs1110400', 'rs11230664', 'rs1805009']	['rs11230664', 'rs1426654', 'rs1448484']
4	['rs1110400', 'rs11230664', 'rs1448484', 'rs1805009']	['rs11230664', 'rs1426654', 'rs1448484', 'rs6119471']
5	['rs1110400', 'rs11230664', 'rs12913832', 'rs1448484', 'rs1805009']	['rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471']

	'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']
--	---	---

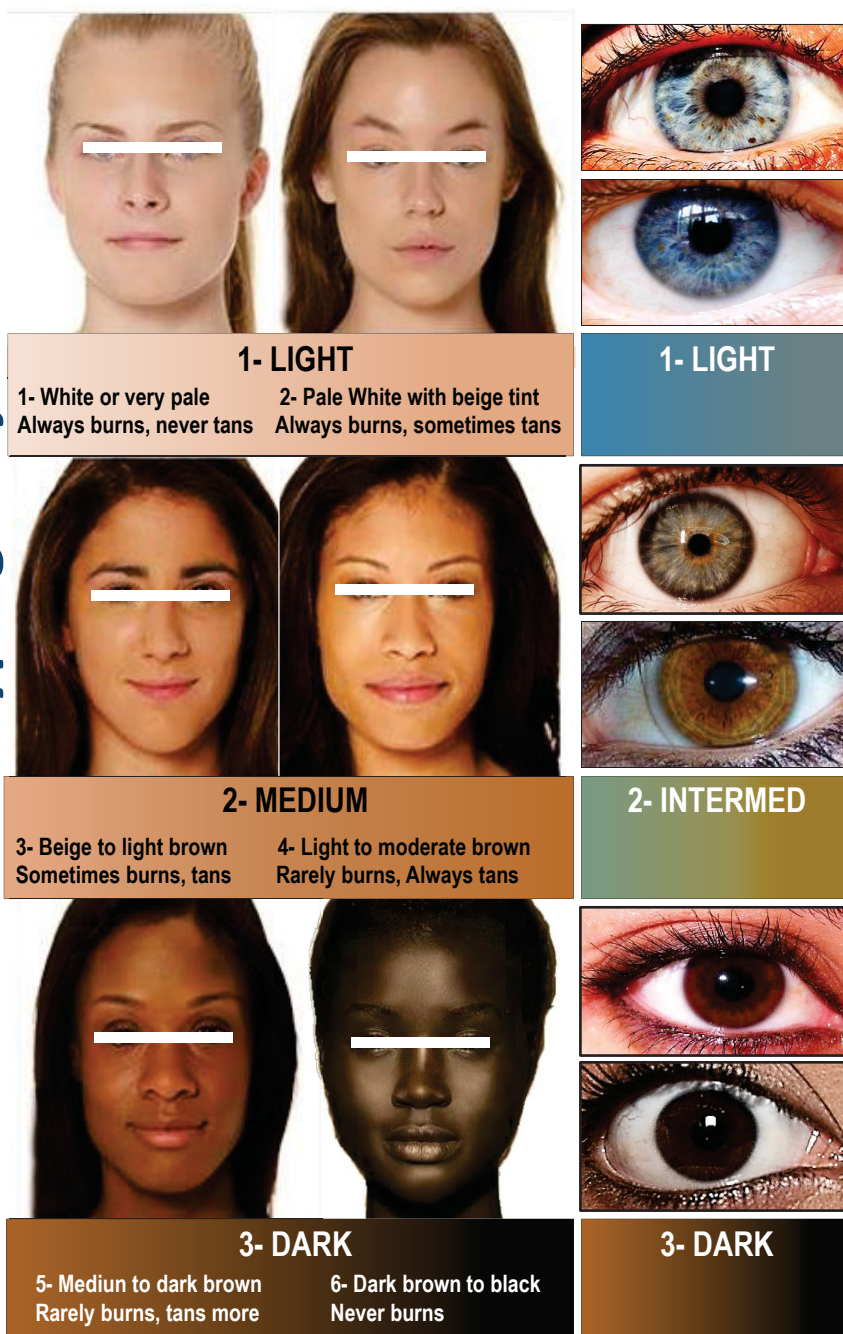
TABLE S17: SNP SELECTION BY RFE AND X² – SKIN 6 CLASSES

SNPs	RFE	X2
2	['rs1110400', 'rs11230664']	['rs11230664', 'rs6119471']
3	['rs1110400', 'rs11230664', 'rs1448484']	['rs11230664', 'rs1426654', 'rs6119471']
4	['rs1110400', 'rs11230664', 'rs12913832', 'rs1448484']	['rs11230664', 'rs1426654', 'rs1448484', 'rs6119471']
5	['rs1110400', 'rs11230664', 'rs12913832', 'rs1448484', 'rs6510760']	['rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471']
6	['rs1042602', 'rs1110400', 'rs11230664', 'rs12913832', 'rs1448484', 'rs6510760']	['rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271']
7	['rs1042602', 'rs1110400', 'rs11230664', 'rs12913832', 'rs1448484', 'rs1805006', 'rs6510760']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271']
8	['rs10424065', 'rs1042602', 'rs1110400', 'rs11230664', 'rs12913832', 'rs1448484', 'rs1805006', 'rs6510760']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271', 'rs8039195']
9	['rs10424065', 'rs1042602', 'rs1110400', 'rs11230664', 'rs12203592', 'rs12913832', 'rs1448484', 'rs1805006', 'rs6510760']	['rs10424065', 'rs10777129', 'rs11230664', 'rs12913832', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271', 'rs8039195']
10	['rs10424065', 'rs1042602', 'rs1110400', 'rs11230664', 'rs12203592', 'rs12913832', 'rs1448484', 'rs1805006', 'rs6119471', 'rs6510760']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1129038', 'rs12913832', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271', 'rs8039195']
11	['rs10424065', 'rs1042602', 'rs1110400', 'rs11230664', 'rs12203592', 'rs12913832', 'rs1448484', 'rs16891982', 'rs1805006', 'rs6119471', 'rs6510760']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1129038', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1448484', 'rs6119471', 'rs6497271', 'rs8039195']
12	['rs10424065', 'rs1042602', 'rs1110400', 'rs11230664', 'rs12203592', 'rs12913832', 'rs1448484', 'rs16891982', 'rs1805006', 'rs6119471', 'rs6510760', 'rs885479']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1129038', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1448484', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195']
13	['rs10424065', 'rs1042602', 'rs1110400', 'rs11230664', 'rs12203592', 'rs12913832', 'rs1448484', 'rs16891982', 'rs1805006', 'rs1805009', 'rs6119471', 'rs6510760', 'rs885479']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1129038', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1448484', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195', 'rs895829']
14	['rs10424065', 'rs1042602', 'rs1110400', 'rs11230664', 'rs12203592', 'rs12913832', 'rs1426654', 'rs1448484', 'rs16891982', 'rs1805006', 'rs1805009', 'rs6119471', 'rs6510760', 'rs885479']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1129038', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1448484', 'rs3794606', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195', 'rs895829']
15	['rs10424065', 'rs1042602', 'rs1110400', 'rs11230664', 'rs1129038', 'rs12203592', 'rs12913832', 'rs1426654', 'rs1448484', 'rs16891982', 'rs1805006', 'rs1805009', 'rs6119471', 'rs6510760', 'rs885479']	['rs10424065', 'rs10777129', 'rs11230664', 'rs1129038', 'rs12913832', 'rs1375164', 'rs1426654', 'rs1448484', 'rs1597196', 'rs3794606', 'rs6119471', 'rs642742', 'rs6497271', 'rs8039195', 'rs895829']

	'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']
67	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']	['rs1037208', 'rs10424065', 'rs1042602', 'rs10756819', 'rs10777129', 'rs1110400', 'rs11230664', 'rs1129038', 'rs11636232', 'rs12203592', 'rs12821256', 'rs12896399', 'rs12913832', 'rs1325127', 'rs13289', 'rs13289810', 'rs1375164', 'rs1393350', 'rs1426654', 'rs1448484', 'rs1597196', 'rs16891982', 'rs16950987', 'rs1724630', 'rs1800404', 'rs1800407', 'rs1805005', 'rs1805006', 'rs1805009', 'rs183671', 'rs1900758', 'rs2036213', 'rs2070959', 'rs2238289', 'rs2240203', 'rs2378249', 'rs2402130', 'rs2424984', 'rs2594935', 'rs2733832', 'rs2835630', 'rs28777', 'rs3212345', 'rs3768056', 'rs3794606', 'rs3935591', 'rs4778137', 'rs4778138', 'rs4778232', 'rs4778241', 'rs4932620', 'rs4959270', 'rs6119471', 'rs642742', 'rs6497271', 'rs6510760', 'rs683', 'rs7170852', 'rs7170989', 'rs7494942', 'rs7948623', 'rs8039195', 'rs885479', 'rs895828', 'rs895829', 'rs916977', 'rs9894429']

ANEXO III - MATERIAIS INFORMATIVOS DO PROJETO

Forensic DNA Phenotyping Project – BRAZIL



(Dr. Alho +55 51 33534334; csalho@pucrs.br; The images are merely illustrative)

Forensic DNA Phenotyping Project – BRAZIL

(Dr. Alho +55 51 33534334; csalho@puccs.br; The images are merely illustrative)



Forensic DNA Phenotyping Project – BRAZIL

	5-BLACK	4-BROWN	3-HAZEL	2-GREEN	1-BLUE

I	White or very pale	Always burns, never tans
II	Pale white with beige tint	Always burns, sometimes tans
III	Beige to light brown	Sometimes burns, always tans
IV	Light to moderate brown	Rarely burns, always tans
V	Medium to dark brown	Rarely burns, tans more than average
VI	Dark brown to black	Never burns

PASSO 1: CUMPRIMENTO, APRESENTAÇÃO E EXPLICAÇÕES



PASSO 2: TCLE



PASSO 3: DADOS

[illegible]

PASSO 5: REGISTRO DAS CARACTERÍSTICAS FÍSICAS



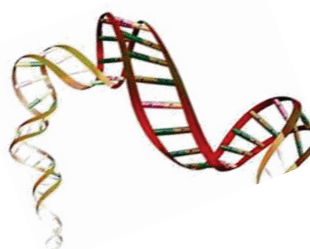
PASSO 4: FOTOS



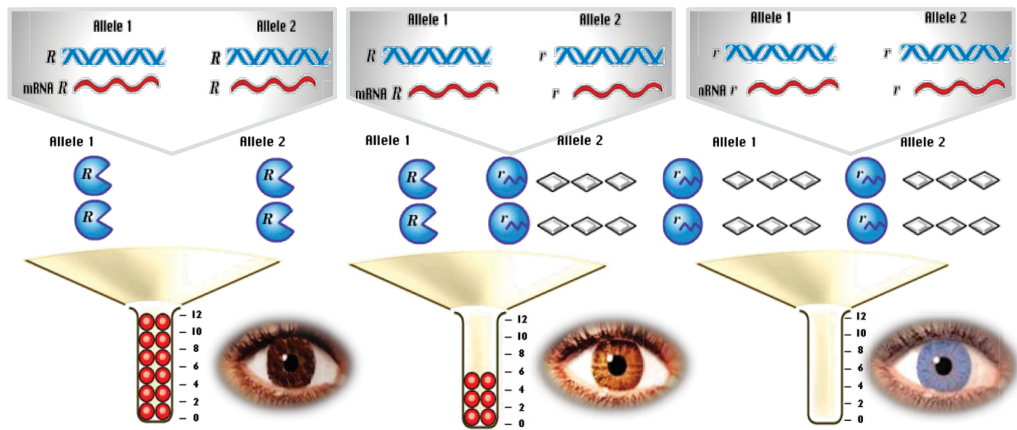
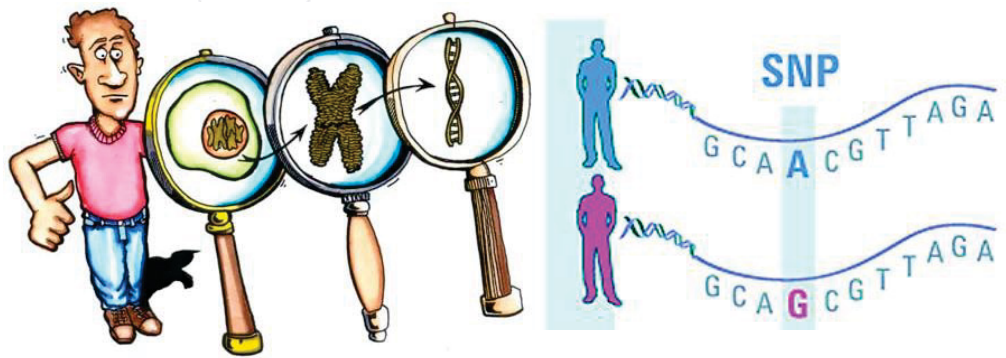
PASSO 6: COLETA DE CÉLULAS



PASSO 7: OBTENÇÃO E ANÁLISE DE DNA



FINAL: COMPARAÇÃO ENTRE O DNA E AS CARACTERÍSTICAS FÍSICAS.



The images are merely illustrative

ANEXO IV – FICHA DE CADASTRO

FORENSIC DNA PHENOTYPING - BRAZIL | INSTRUMENTO DE DADOS

DATA DA COLETA: ____ / ____ / ____

IDENTIFICAÇÃO:

NUMERO				INICIAIS	

DADOS PESSOAIS DO PARTICIPANTE

Nome: _____ Data de Nascimento: ____ / ____ / ____

Contato (E-mail; telefones): _____

DADOS ANTROPOMÉTRICOS / CARACTERÍSTICAS / INFORMAÇÕES

Altura: _____ Peso: _____ Nº do sapato: _____

Ao usar a mão para escrever é: () Destro () Canhoto

A orelha é saliente? () SIM () NÃO Lóbulo orelha é solto? () SIM () NÃO

Está usando lente de contato? () SIM () NÃO Está bem bronzeado? () SIM () NÃO

Cabelo pintado/descolorido...? () SIM () NÃO ...Alisado/Encrespado? () SIM () NÃO

DADOS EPIDEMIOLÓGICOS

Sexo: () Masc () Fem

Origem: () Europeia () Africana () Indígena () Asiática () Árabe* () Outra: _____

Residente em: () Sul () Sudeste () Centroeste () Nordeste () Norte () Estado: _____

Natural de: () Sul () Sudeste () Centroeste () Nordeste () Norte () Estado: _____

(*Árabes: Habitantes da Península Arábica, Oriente Médio, África setentrional).

DADOS BIOLÓGICOS - ORIGINAIS (NATURAIS)

OLHO COR: () Azul () Verde () Mel () Marrom () Preto

PELE COR: () I-Leite () II-Bege () III-Cuia () IV-Chocolate () V-Café () VI-Petróleo

CABELO COR: () Loiro Plat () Loiro Amar () Loiro Esc () Castanho Claro () Cast Esc () Negro

() Ruivo Claro () Ruivo Esc

CABELO TIPO: () Liso () Ondulado () Crespo () Molinha

CABELO QTDE: () Muito () Médio () Pouco CABELO OBSERV.: () Calvície () Grisalho

() Branco () _____

PARA HOMENS

QTD DE PELO EM FACE/PEITO/COSTAS/PERNAS/BRAÇOS: () Muito () Médio () Pouco

COLETADOR

Nome COLETADOR: _____	Matrícula PUCRS: _____
Como coletador, declaro que colhi e identifiquei corretamente todas as amostras de células da pessoa aqui indicada para fins de investigação de dados fenotípicos, tendo seguido rigorosamente os procedimentos indicados. Para tanto assino ao lado:	
Assinatura COLETADOR _____	

OBSERVAÇÕES

(SE FOR O CASO: DESENHE O HEREDOGRAMA DA FAMÍLIA À QUAL O PARTICIPANTE PERTENCE, IDENTIFICANDO-O COM UMA SETA).



ANEXO V – TERMO DE CONSENTIMENTO

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE) NO DESENVOLVIMENTO DA PRESENTE PESQUISA REFERENTE AO ESTUDO DE FENÓTIPO (POPULAÇÃO 2)

Você está sendo convidado para participar de uma pesquisa. O título da pesquisa é MARCADORES GENÉTICOS PARA AS CARACTERÍSTICAS EXTERNAS VISÍVEIS, e tem por objetivo estudar partes do material genético das pessoas que possam estar relacionadas a características de cor dos olhos/pele/cabelo, tipo e quantidade de pelo/cabelo, formato do rosto, peso, altura. Para isso precisamos: 1- anotar estes dados em um formulário; 2- fotografar seu rosto de frente e dos lados (umas 10-12 fotos); 3- coletar células da parte interna da sua boca com um cotonete esterilizado. Talvez a coleta da saliva lhe cause um desconforto momentâneo, mas nós qualificamos que sua participação nesta pesquisa lhe oferece um risco mínimo. Você não terá custo algum em participar deste estudo. Este estudo não trará benefícios para você diretamente, nesse momento. Mas, haverá benefícios para a sociedade e para a ciência se conseguirmos encontrar os resultados que buscamos; isso poderá ajudar a identificar pessoas desaparecidas, por exemplo.

Sinta-se à vontade para fazer qualquer pergunta sobre o estudo e esclarecer suas dúvidas. Os responsáveis pela pesquisa garantidamente lhe responderão. Caso você tenha vontade de ser retirado da pesquisa, diga isso a qualquer momento, sem necessidade de se explicar; você não será prejudicado se resolver tomar essa decisão. Apenas comunique aos pesquisadores se quiser abandonar a pesquisa e sua decisão será respeitada. Toda e qualquer informação atualizada sobre este estudo estará à disposição de todos os que participam da pesquisa. Os resultados detalhados deste estudo poderão ser comunicados a você, sempre que você quiser. Sua dignidade e privacidade, em relação a qualquer dado, serão sempre mantidas. Seus dados estarão protegidos contra qualquer tipo de uso que não os aqui previstos; os dados receberão um número para garantir seu anonimato, ou seja, no estudo dos dados, a sua identidade será desconhecida tanto pela equipe do trabalho como por qualquer outra pessoa externa. Todos os documentos desta pesquisa, como o projeto (com base teórica, hipótese, objetivo, metodologia, cronograma, orçamento), e este termo, foram avaliados e aprovados por um Comitê de Ética em Pesquisa (CEP) vinculado à Comissão Nacional de Ética em Pesquisa (CONEP). CEP é um grupo independente de examinadores de diversas áreas (biológicas, de saúde, exatas, sociais e humanas) que avalia os aspectos ético-metodológicos de pesquisas feitas com pessoas. É uma avaliação bem criteriosa e longa, e serve para garantir que tudo na pesquisa seja feito de forma correta. Assim, nós gostaríamos de pedir seu consentimento para participar deste estudo, e também para que o material biológico que você forneceu seja guardado de forma segura para ser usado em pesquisas futuras, desde que elas também tenham sido autorizadas pelo CEP/CONEP.

Eu, (preencher com nome do participante) fui informado dos objetivos da pesquisa de maneira clara e objetiva. Recebi informação a respeito dos procedimentos do estudo e esclareci minhas dúvidas. Sei que em qualquer momento poderei solicitar novas informações e/ou modificar minha decisão de participar. A equipe da Dra. Clarice S. Alho certificou-me de que todos os meus dados serão confidenciais, e minha rotina de vida não será modificada em razão desta pesquisa. Acredito que fui suficientemente informado a respeito das informações que estão neste termo, e concordo voluntariamente em participar deste estudo. Por isso, deixo aqui meu consentimento. Caso eu tenha novas perguntas sobre a pesquisa, posso telefonar para a Dra. Clarice S. Alho no telefone (51) 3320 3545, ramal 4534, ou para esclarecer sobre os meus direitos como participante deste estudo, ou se penso que fui prejudicado pela minha participação, posso chamar também os responsáveis pelo Comitê de Ética em Pesquisa da PUCRS (CEP-PUCRS) no telefone (51) 3320 3000 ramal 3345. Entendi que o material e os dados obtidos a partir desta pesquisa poderão ser usados em pesquisas futuras, feitas aqui ou fora do país, mas isso acontecerá se elas também forem avaliadas e aprovadas pelo CEP/CONEP. Se uma nova pesquisa aprovada pelo CEP/CONEP for realizada, eu:

() dispenso que me apresentem um novo termo para assinar.


() peço que me apresentem um novo termo para assinar.

E-mail / telefones: _____

Nome e assinatura do participante	Local e data
Nome e assinatura do pesquisador	Local e data

- ✓ O CEP-PUCRS tem como endereço: Avenida Ipiranga, 6690 - Prédio 40 - 5º andar – Sala 505 – Porto Alegre/RS; telefone (51) 3320 3000 ramal 3345.
- ✓ Este Termo de Consentimento foi elaborado em DUAS VIAS, sendo uma ficará com você e a outra com o pesquisador, conforme a Resolução CNS 466/2012 itens IV.3.f, IV.5.d.

ANEXO VI – PARECER DO COMITÊ DE ÉTICA

<p>PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL - PUC/RS</p> 	
PARECER CONSUBSTANCIADO DO CEP	
DADOS DO PROJETO DE PESQUISA	
Título da Pesquisa: FREQUÊNCIAS ALÉLICAS E GENOTÍPICAS DE SNPs EM GENES QUE REGULAM A PIGMENTAÇÃO DE OLHOS PARA FINS FORENSES	
Pesquisador: CLARICE SAMPAIO ALHO	
Área Temática: Genética Humana: (Haverá envio para o exterior de material genético ou qualquer material biológico humano para obtenção de material genético, salvo nos casos em que houver cooperação com o Governo Brasileiro;);	
Versão: 1	
CAAE: 23034713.7.0000.5336	
Instituição Proponente: UNIAO BRASILEIRA DE EDUCACAO E ASSISTENCIA	
Patrocinador Principal: Financiamento Próprio	
DADOS DO PARECER	
Número do Parecer: 508.482	
Data da Relatoria: 20/12/2013	
<p>O projeto esta muito bem elaborado e não existem pendencias ou inadequações.</p>	
Situação do Parecer: Aprovado	
Necessita Apreciação da CONEP: Sim	
Considerações Finais a critério do CEP: Ratificamos o parecer do relator.	
O presente projeto, seguiu nesta data para análise da CONEP e só tem o seu início autorizado após a aprovação pela mesma.	
<p>PORTO ALEGRE, 09 de Janeiro de 2014</p>	
<p>Assinador por: João Feliz Duarte de Moraes (Coordenador)</p>	